THE PRAXEOLOGY OF PRIVACY

PRAISE FOR THE PRAXEOLOGY OF PRIVACY

"The Praxeology of Privacy does an excellent job integrating the Axiom of Resistance into a comprehensive praxeological treatment of privacy as selective disclosure and strategic defense. In weaving together Austrian action principles, Cypherpunk ethos, and software engineering, it offers a practical roadmap for building parallel institutions that raise the cost of surveillance. The synthesis is both coherent and timely, showing why resistant money, anonymous communication, and decentralized protocols matter well beyond any single technology. The core synthesis pertaining to action and resistance is powerful, and readers will gain a great deal from the book's rigorous integration of theory and implementation. Highly recommended for anyone serious about building a freer future."

— **Eric Voskuil**, author of *Cryptoeconomics*

"Max Hillebrand's book recommendations introduced me to two of my favorite books of all time. His own book is now added to that list. The Praxeology of Privacy manages to seamlessly apply Misesian logic to the Internet world we live in today, pointing out the dangers of the sunken cost of mass surveillance, and why the future still looks bright because of all the innovation in Freedom-go-Up technology that Max has dedicated his life to help manifest into reality."

— **Knut Svanholm**, author of *Praxeology: The Invisible Hand That Feeds You*

"Max Hillebrand's The Praxeology of Privacy builds on the thought of Mises, Rothbard, and Hoppe, as well as that of other thinkers, such as myself, to continue to extend the science of liberty to further domains and topics. Hillebrand dispatches with the notion of an independent 'right to privacy' not anchored in property rights in material resources. Privacy is not a separate property right; it's what results when self-ownership and property rights are respected. Nonetheless, privacy is crucially important: it is the ability to selectively reveal oneself to the world, an aspect of purposeful behavior. Hillebrand also develops the axiom of resistance, based on Voskuil's Cryptoeconomics, to explain how rights to control resources require the ability to resist external control. This is a fascinating and ambitious work, sure to delight and provoke those hungry for further exploration of praxeology and liberty."

— **Stephan Kinsella**, author of *Against Intellectual Property*

"Max has been working on and advocating for privacy for as long as I've known him, and The Praxeology of Privacy is the book I've been waiting for this whole time. This book bridges the cypherpunk ethos and Austrian economics, helping people from each tradition understand the other. In the modern era of state surveillance and control, The Praxeology of Privacy is a practical guide to defending your privacy."

— **Luke de Wolf**, author of *Defending Bitcoin*

THE PRAXEOLOGY OF PRIVACY

Economic Logic in Cypherpunk Implementation

Third Edition --- Extended

MAX HILLEBRAND

The Praxeology of Privacy
Economic Logic in Cypherpunk Implementation

Third Edition — Extended Edition

Digital / e-reader edition with full annotated endnotes.

Max Hillebrand

This work is released into the **Public Domain**.

No rights reserved. You are free to copy, modify, distribute, and perform this work, even for commercial purposes, without asking permission.

FOREWORD

The Praxeology of Privacy is an extraordinary book, and the chief reason for me saying so is simple: I spent twenty years in the privacy trenches, and during that time I learned many lessons, at the cost of great time and effort. In this book, however, Max Hillebrand has gathered all of those lessons, and more beside, into a book that you can obtain for a modest cost, and from which you can deeply educate yourself in a fairly small number of hours.

I would have paid dearly for such easy to access information years ago.

There are, to me, three prominent benefits to this book. The first is simply that Max explains not just the things themselves, but why they are true and why they operate as they do. If you're at all like me, this is the ingredient that holds the entire discourse together. Lists of facts are fine, but an understanding of how and why they operate is what makes the whole recognizable and intelligible — and memorable.

The second factor is one that I know from my years of friendship with Max, and which I'm especially pleased to pass along to you: Max has lived this, not just observed it. Max has lived the cypherpunk life. (Cypherpunks being privacy and encryption advocates.) He has had to put the ideas in this book to the test in real life, and for many years now. That inevitably strips the metal from the dross and is essential to any practical understanding of what works and doesn't. Trial in the real world rips away misunderstanding and thin understandings like pretty much nothing else. So, nice-sounding ideas which couldn't survive hostile contact with the world were stripped away before this book was written.

The third reason for my valuing this book is one you might not expect from the title: By the time you're done you'll have received an excellent education in economics. While this is far from an economics text, the overlap between human motivations, incentives and costs (monetary or otherwise) associated with all of this are perhaps best explained in economic terms. And so this best way of explaining exposes you to a great deal of economics along the way, even though it doesn't feel like you're learning economics.

For many years we've lacked this level of examination of privacy in human life and human affairs. Now we have one, and I highly recommend that you get a copy. You'll need either this or a career in privacy to grasp the information economics of the 21st century. Please believe me that this way is easier.

Paul Rosenberg
May 2026

PREFACE

Privacy is the precondition of every voluntary arrangement a free person enters into, and it is being engineered out of daily life. This book is about the conditions under which it survives. Privacy in the working sense developed here is selective disclosure: the power to choose what is revealed and to whom. Selective disclosure is a stronger notion than concealment or secrecy, and conflating the three is the main reason public debate about surveillance stays stuck.

The subject has become urgent because two infrastructures matured at the same moment. Cryptographic capability sufficient to protect ordinary communication, payments, identity, and computation is now in production: post-quantum signatures, threshold systems, zero-knowledge proofs, and computation on encrypted data are deployed and used. So is the observation stack assembled on the other side. Commercial spyware reaches opposition journalists, lawyers, and politicians through the phones in their pockets.[1] Data brokers sell aggregated records to the agencies that once had to obtain them directly. Proposals in major jurisdictions would mandate scanning inside encrypted applications, removing the protection end-to-end encryption provides.[2] Programmable money with spending restrictions defined by the issuer is moving from research into pilots.[3] Two architectures are contending for default status. Whichever becomes common infrastructure will be difficult to replace inside the ordinary political calendar.

The pressure is not confined to any one population. A journalist protecting a source, a small business owner whose payment processor froze an account without explanation, a doctor whose patient records sit on a vendor's server, a parent whose child's school issues mandatory tracked devices, an opposition organizer in a country with hostile authorities, a developer who would like

to ship a working system without an enforcement letter: all are operating inside the same observation environment, with the same primitives available to defend themselves. The question of who can audit, who can refuse, and who can build is no longer specialist.

This book joins two traditions that arrived at the same conclusions from different starting points. Austrian economics, through deduction from the fact that human beings act, established that privacy is structural to deliberation and exchange, that sound money is essential to economic coordination, that the state is systematic aggression on whatever scale its enforcement capacity permits, and that the socialist calculation problem makes central planning incoherent.[4] The cypherpunk tradition, through cryptographic engineering grounded in agorist and voluntaryist thought,[5] established that privacy can be defended in production, that sound money can be programmed, that systems can be built to resist control at a cost their operators can afford, and that the infrastructure can be open enough for any user to audit. Their synthesis is the purpose of this book.

An old and practical gap runs between them. Austrian writing rarely commits to the implementation question, and much of it stops at diagnosis; a state threatened by software will not be persuaded by another essay. Cypherpunk practice often commits to implementation before it has settled the political-economy question; it builds working cryptography that can still fail as a social system. Builders centralize what should remain distributed, make compromises that betray the purpose of the system, lose to incentives they never audited, and ship products that reverse the protection their technology provides. Theorists publish explanations of the trap and leave the tools to others. Each tradition needs the other, and what they share is enough to rebuild the defensive layer the surveillance era has dismantled.

The book is a treatise: a sustained argument that proceeds by definition, derivation, scope fence, and cross-reference. Each chapter states what it establishes, derives the conclusion from prior chapters, fences its scope against overreach, and points forward to where the derivation continues. Three axioms do the structural work. Mises's action axiom treats privacy as a condition of purposive action to-

ward chosen ends; Hoppe's argumentation axiom treats privacy as a condition of rational discourse; Voskuil's axiom of resistance treats privacy as a consequence of raising the cost of observation above the observer's willingness to pay. The chapters operationalize these axioms against the cryptographic primitives that make the engineering possible.

The Austrian line this book inherits runs through Mises[6] on action, Rothbard[7] on property, Hoppe[8] on argumentation, Hayek[9] on the knowledge problem, and Oppenheimer[10] on the state's origin. For the cypherpunk line: May[11] and Hughes[12] on the political meaning of cryptography, Finney[13] on running code, Nakamoto[14] on sound digital money, and Voskuil[15] on the axiom of resistance. Agorist and voluntaryist writing that shaped the cypherpunks' self-understanding appears where it carries weight. The privacy claims the book defends rest on three rights that are common ground across the libertarian-Austrian tradition: self-ownership of the body, property rights in physical resources, and contract. "Privacy rights" as a freestanding category is not invoked; Chapter 6 develops the property-theoretic frame in detail and marks where the live intra-Austrian disagreements over patent and copyright sit.

A reader from the Austrian tradition will find the theory operationalized. Cryptographic primitives translate through economic analogies, and the engineering arguments are formally compatible with the axiomatic method. A reader from the cypherpunk tradition will find the economic and praxeological frame much of the tradition already draws on implicitly: why these systems matter beyond their technical merit, which compromises preserve their purpose and which destroy it, and how the sound-money and cryptographic projects converge into a single operational stack. A reader from neither tradition, who senses that something has tightened and wants to understand why, will find the argument accessible from first principles. The only prerequisite is seriousness.

Privacy is part of the conditions under which human beings deliberate, exchange, save, coordinate, and live unobserved when they choose to. The strategic claim the chapters develop is simple. State predation depends on state observation. When observation

gets cheaper, predation gets cheaper; when observation gets more expensive, predation recedes. The engineering that raises the cost of observation is the engineering this book is about.

Notes

[1]Citizen Lab at the University of Toronto, https://citizenlab.ca/, is the primary empirical source documenting commercial-spyware infections of journalists, dissidents, lawyers, and politicians across more than thirty state customers. Pegasus from NSO Group, https://www.nsogroup.com/, is the reference product; *WhatsApp LLC v. NSO Group*, Case No. 4:19-cv-07123 (N.D. Cal. 2024), produced the first major civil accountability ruling against a spyware vendor.

[2]EU Regulation on Child Sexual Abuse ("Chat Control" / CSAR), proposal COM(2022) 209, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0209; UK Online Safety Act 2023, Section 122, https://www.legislation.gov.uk/ukpga/2023/50/section/122. The European Court of Human Rights held in *Podchasov v. Russia* (App. no. 33696/19), February 13, 2024, https://hudoc.echr.coe.int/, that a statutory requirement to weaken end-to-end encryption "cannot be regarded as necessary in a democratic society."

[3]Atlantic Council, "Central Bank Digital Currency Tracker," https://www.atlanticcouncil.org/cbdctracker/, tracks more than one hundred jurisdictions across research, pilot, and live phases; Human Rights Foundation, "CBDC Tracker," https://cbdctracker.hrf.org/, focuses on the civil-liberties implications.

[4]Ludwig von Mises, "Economic Calculation in the Socialist Commonwealth," trans. S. Adler, in *Collectivist Economic Planning*, ed. F. A. Hayek (London: Routledge, 1935; orig. *Die Wirtschaftsrechnung im sozialistischen Gemeinwesen*, 1920), https://mises.org/library/essay/economic-calculation-socialist-commonwealth. The paper that launched the socialist-calculation debate: rational economic calculation requires private ownership of the means of production and real market prices, and no socialist planner can replicate either. The calculation problem and the knowledge problem (note 7) are complementary: together they establish that central planning is incoherent in principle.

[5]Samuel Edward Konkin III, *New Libertarian Manifesto* (Koman Publishing, 1980; 2nd ed. 1983), https://agorism.info/NewLibertarianManifesto.pdf. The founding document of agorism: the path to a free society runs through counter-economics, peaceful voluntary exchange operating outside the official economy, not through electoral politics. Konkin coined the terms "agorism" and "counter-economics." The voluntaryist tradition, associated with Carl Watner and the journal *The*

Voluntaryist (`https://voluntaryist.com/`), reaches similar conclusions through principled non-participation. Both strands shaped the cypherpunks' conviction that building working systems matters more than winning political arguments.

[6]Ludwig von Mises, *Human Action: A Treatise on Economics*, 4th rev. ed. (Irvington-on-Hudson, NY: Foundation for Economic Education, 1996; orig. 1949), `https://mises.org/library/book/human-action`. The foundational text of praxeology: economics derived by strict deduction from the action axiom, the fact that humans act purposefully to remove felt unease. Mises's account of economic calculation (pp. 200–231) establishes why deliberation requires undisclosed plans; any observation that reveals those plans to adversaries degrades the actor's capacity to achieve his ends. The Mises Institute hosts the full text without charge.

[7]Murray N. Rothbard, *Man, Economy, and State: A Treatise on Economic Principles*, 2nd scholar's ed. (Auburn, AL: Ludwig von Mises Institute, 2004; orig. 1962), `https://mises.org/library/book/man-economy-and-state-power-and-market`; and Rothbard, *The Ethics of Liberty* (New York: New York University Press, 1998; orig. 1982), `https://mises.org/library/book/ethics-liberty`. *Man, Economy, and State* extends Mises's praxeology into price theory, capital, and production structure, grounding all exchange in voluntary consent. *The Ethics of Liberty* grounds individual rights in self-ownership and the non-aggression principle, establishing the property framework the book's privacy argument inherits throughout.

[8]Hans-Hermann Hoppe, *The Economics and Ethics of Private Property: Studies in Political Economy and Philosophy*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2006), `https://mises.org/library/book/economics-and-ethics-private-property`. Hoppe's argumentation ethics derives property rights from the performative presuppositions of rational discourse: anyone who argues must already presuppose self-ownership and control of the means by which they argue, since denying that control is self-refuting. Chapter 4 develops this framework as the normative basis for treating privacy as a condition of rational discourse.

[9]Friedrich A. Hayek, "The Use of Knowledge in Society," *American Economic Review* 35, no. 4 (1945): 519–530, `https://www.econlib.org/library/Essays/hykKnw.html`. The canonical statement of the knowledge problem: prices coordinate dispersed local knowledge that no central authority can collect or process, and any system suppressing the price mechanism also suppresses the information it carries. Hayek, *The Constitution of Liberty* (Chicago: University of Chicago Press, 1960), Part II, develops the related argument that freedom is partly defined by the space where one's plans need not be disclosed to authorities.

[10]Franz Oppenheimer, *The State: Its History and Development Viewed Sociologically*, trans. John M. Gitterman (New York: B. W. Huebsch, 1914; orig. *Der Staat*, 1908), `https://mises.org/library/book/state`. Oppenheimer's canonical distinction between the "economic means" (voluntary production and exchange)

and the "political means" (coercive expropriation) establishes the analytical basis for treating the state as predatory by origin. Albert Jay Nock, *Our Enemy, the State* (Caldwell, ID: Caxton Printers, 1935), applies this framework to American history. Both works underlie Chapter 4's treatment of state surveillance as predation.

[11]Timothy C. May, "The Crypto Anarchist Manifesto" (1988), `https://www.activism.net/cypherpunk/crypto-anarchy.html`; and Timothy C. May, "Crypto Anarchy and Virtual Communities" (1994), `https://nakamotoinstitute.org/library/virtual-communities/`. May predicted that public-key cryptography would produce irreversible changes in the scope of government regulation, the ability to tax and control economic interactions, and the way trust and reputation function, formulating the political program before the infrastructure existed to execute it.

[12]Eric Hughes, "A Cypherpunk's Manifesto" (March 9, 1993), `https://www.activism.net/cypherpunk/manifesto.html`. The operational complement to May's manifesto: where May identifies the political stakes, Hughes specifies the engineering requirement. Privacy must be defended by anonymous transaction systems and open cryptographic code, not by trusting institutions. Hughes's definition of privacy as "the power to selectively reveal oneself to the world" is the definition this book adopts.

[13]Hal Finney, "RPOW: Reusable Proofs of Work" (August 15, 2004), `https://nakamotoinstitute.org/rpow/`; and Finney, "Bitcoin and me," BitcoinTalk (March 19, 2013), `https://bitcointalk.org/index.php?topic=155054.0`. Finney's RPOW system was the direct technical predecessor of Bitcoin's mining mechanism and the first working implementation of a transferable digital token backed by computational work. He was also the recipient of the first Bitcoin transaction (from Nakamoto, January 12, 2009), and his account of early involvement describes Bitcoin as the culmination of the cypherpunk project.

[14]Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (October 31, 2008), `https://bitcoin.org/bitcoin.pdf`. The nine-page whitepaper introducing Bitcoin: a decentralized electronic cash system using a proof-of-work chain to achieve consensus without a trusted third party. The privacy model in Section 10 is built on pseudonymity and key separation. The whitepaper is the founding document of the cryptocurrency field and the first working demonstration that sound digital money can be engineered without a central issuer.

[15]Eric Voskuil, *Cryptoeconomics: Fundamental Principles of Bitcoin* (Libbitcoin Institute, 2020), `https://github.com/libbitcoin/libbitcoin-system/wiki/Cryptoeconomics`. Voskuil's axiom of resistance holds that a Bitcoin security model presupposes people willing to defy state coercion at personal cost; without this assumption, Bitcoin reduces to a permissioned database. The axiom is the third structural premise this book employs alongside Mises's action axiom and Hoppe's argumentation axiom.

CONTENTS

Foreword . . . . . **vii**

Preface . . . . . **ix**

I Foundations **1**

1 The Nature of Privacy **3**

1.1 Privacy as Control Over Disclosure 3

1.2 Three Related Terms 4

1.3 Privacy as Strategic Defense 6

1.4 The Book's Argument in Brief 6

2 Two Traditions, One Conclusion **13**

2.1 The Austrian Approach: Deduction from Action . . . 14

2.2 The Cypherpunk Approach: Building What Theory Requires 17

2.3 Independent Convergence: Why Both Reached the Same Place 21

II Axioms **27**

3 The Action Axiom: Privacy as Structural Feature **29**

3.1 The Action Axiom 30

3.2 Internal Deliberation and Subjective Valuation 32

3.3 Information Asymmetry and Control 34

4 The Argumentation Axiom and Self-Ownership 39
4.1 The Argument: Performative Contradiction 40
4.2 Objections and Responses 42
4.3 The Is-Ought Question 48
4.4 Implications: Property and Non-Aggression 50
4.5 The Argumentation Axiom and Privacy 51
5 The Axiom of Resistance 59
5.1 What the Axiom Asserts 60
5.2 Why the Axiom Is Well-Grounded 62
5.3 Relationship to Other Foundations 65

III Economic Foundations 69

6 Information, Scarcity, and Property 71
6.1 Scarcity as the Foundation of Property 72
6.2 "Intellectual Property" as Aggression 74
6.3 Content vs. Media: The Critical Distinction 76
6.4 How Privacy Is Protected 77
6.5 Implications for Privacy Analysis 80
7 Exchange Theory and Privacy 91
7.1 The Logic of Exchange 91
7.2 How Privacy Enhances Exchange 93
7.3 How Surveillance Distorts Exchange 95
7.4 Exchange Can Occur Under Surveillance 97
7.5 Specific Exchange Contexts 98

8 Capital Theory and Entrepreneurship 107
8.1 Privacy Infrastructure as Capital Goods 108
8.2 Time Preference and the Privacy-Convenience Tradeoff 109
8.3 Entrepreneurial Discovery in Privacy 111
8.4 Capital Heterogeneity and Privacy Specialization . . 113
8.5 Capital Formation and Privacy Development 114
9 Monetary Theory and Sound Money 121
9.1 The Market Origin of Money 122
9.2 The Regression Theorem 123
9.3 Sound Money Properties 125
9.4 Fiat Money and Its Problems 126
9.5 Free Banking vs. Central Banking 128
9.6 Money Proper vs. Money Substitutes 129
9.7 Digital Money Requirements 130
9.8 Bridge to Bitcoin 131

IV The Adversary 137

10 Financial Surveillance and State Control 139
10.1 Intervention Theory: Autistic, Binary, Triangular . . . 139
10.2 Autistic Intervention and Privacy 140
10.3 Binary Intervention and Privacy 141
10.4 Triangular Intervention and Privacy 142
10.5 Central Bank Digital Currencies as Total Intervention 146
10.6 The Adversarial Decision Cycle 154

11 Corporate Surveillance and Data Extraction 165
11.1 The Business Model of Data Extraction 165
11.2 Corporate-State Entanglement 168
11.3 Is This a Market Failure? 170
11.4 The Irreversibility Problem 173
11.5 Market Responses and Privacy Competition 176
12 The Analytics Stack 187
12.1 The Division of Labor in Surveillance 187
12.2 Machine Vision at Scale 189
12.3 The Model as Analyst 190
12.4 The Data-Broker Channel 191
12.5 Physical-World Infrastructure 193
12.6 The Commercial Spyware Arm 195
12.7 Compressing the OODA Loop 197
12.8 Why Agencies Keep Buying 199
12.9 Why the Stack Exists Now 200
13 The Crypto Wars 209
13.1 History: Export Controls to Clipper Chip 209
13.2 The Economic Logic of Cryptographic Control 212
13.3 Why Control Fails (and Where It Doesn't) 214
13.4 Jurisdictional Competition and Arbitrage 216
13.5 The Ongoing War 217

V Technical Implementation 231
14 Cryptographic Foundations 233
14.1 Symmetric and Asymmetric Cryptography 234
14.2 Hash Functions and Digital Signatures 237
14.3 Trust: Mathematical vs. Institutional 239
14.4 Limitations and Vulnerabilities 241

15 Zero-Knowledge Proofs 251
15.1 The Verification Dilemma 252
15.2 The ZK Concept: Proving Without Revealing 253
15.3 Types: SNARKs, STARKs, Bulletproofs 254
15.4 Applications: Zcash, Rollups, Identity 256
15.5 Economic Implications: Verification as Service 262
15.6 Limitations 263
16 Computing on Secrets 271
16.1 The Verification-Execution Distinction 272
16.2 Homomorphic Encryption: Arithmetic on Ciphertext 273
16.3 Secure Multi-Party Computation 275
16.4 Trusted Execution Environments 278
16.5 Private Information Retrieval 283
16.6 Differential Privacy and the Aggregation Layer 285
16.7 The Design Space and the Complements 288
16.8 The Axiom of Resistance Applied to Compute 289
17 Anonymous Communication Networks 299
17.1 The Problem: How the Internet Leaks Privacy 300
17.2 Requirements for Anonymous Communication 301
17.3 VPNs: A Simple but Limited Solution 304
17.4 Onion Routing: Tor and I2P 306
17.5 Mixnets: The Strongest Protection 311
17.6 Why No Single Anonymity Architecture Suffices . . . 313
17.7 When the Internet Is Not the Network 314
17.8 Encrypted Messaging: The Application Layer 317
18 Bitcoin and the Digital Money Breakthrough 333
18.1 The Twin Problems of Digital Money 333
18.2 Precursors and Their Failures 335
18.3 Nakamoto Consensus 337

19 Bitcoin as Sound and Resistant Money 351
19.1 Sound Money Properties as Schelling Points 351
19.2 Resistance Properties: Why Bitcoin Survives 358
20 Bitcoin Privacy and Monetary Layers 369
20.1 How Base-Layer Privacy Fails 370
20.2 Programmable Money: Script Primitives 379
20.3 Privacy Solutions: Collaborative Transactions and Swaps 381
20.4 Non-Custodial Off-Chain Payments 386
20.5 Custodial Privacy Layers 393
20.6 Shielded Client-Side Validation 397
20.7 Privacy Coin Comparators 400
20.8 The Quantum Horizon for Bitcoin 411
21 Decentralized Social Infrastructure 429
21.1 The Problem with Centralized Platforms 429
21.2 Nostr: The Protocol Solution 433
21.3 Keys as Identity: Cryptographic Sovereignty 435
21.4 Relay Architecture and Market Dynamics 438
21.5 Reputation Without Central Authority 440
21.6 Moderation as Market Service 443
21.7 Beyond Social Media: One Protocol, Many Payloads 444
21.8 Privacy Limitations 445
21.9 Why Alternatives Fall Short 451

VI Praxis **463**

22 Operational Security **465**

22.1 Threat Modeling: Who Is Your Adversary? 465

22.2 The Weakest Link: Human Factors 469

22.3 Technical Security Fundamentals 472

22.4 Compartmentalization and Identity Separation 477

22.5 Organizational Security and the Infiltration Problem 480

22.6 Venue Security and Physical Nodes 482

22.7 Surveillance Detection 483

22.8 Common Failure Modes (Case Studies) 485

22.9 The Limits of OPSEC 487

23 Implementation Strategy **497**

23.1 Starting Where You Are 497

23.2 Before You Begin: Apply Your Threat Model 499

23.3 Progressive Steps: From Simple to Advanced 500

23.4 Finding Your Community 507

23.5 From Individual Practice to Shared Practice 510

23.6 What Not to Do 512

24 Trust and Dispute in the Parallel Economy **525**

24.1 Why Tools Alone Do Not Make a Market 526

24.2 Reputation Under Pseudonymity 527

24.3 Credentials and Selective Disclosure 529

24.4 Escrow and Simple Dispute Resolution 532

24.5 Arbitration and Contested Facts 533

24.6 Merchant Discovery and Institutional Emergence . . 535

24.7 Present Limits and Honest Boundaries 536

25 Building the Parallel Economy 545
25.1 The Convergence of Foundations 545
25.2 The Synthesis 546
25.3 Breaking the Observation Loop 547
25.4 What Has Been Achieved 548
25.5 Limits and Open Questions 553
25.6 Answering "Nothing to Hide" 558
25.7 Conclusion 560
About This Edition 565

PART I

FOUNDATIONS

1

THE NATURE OF PRIVACY

> "Privacy is necessary for an open society in the electronic age."
>
> Eric Hughes, *A Cypherpunk's Manifesto* (1993)[1]

"If you have nothing to hide, you have nothing to fear."

That line appears wherever privacy is discussed, and it works because it smuggles in a false definition before the argument even begins.[2] It redefines privacy as concealment. Once that move succeeds, resistance to surveillance can be painted as suspicious from the start.

The right response is not embarrassment, nor the weak claim that everyone has private shame he would prefer to keep off display. The right response is definitional: privacy is not the hiding of guilt but control over disclosure. Once that point is fixed, the whole argument changes.

The full answer to "nothing to hide" comes at the end of the book in Chapter 25. For now, the task is to make the terms clear enough that the argument cannot cheat.

1.1 Privacy as Control Over Disclosure

Arguments about privacy usually collapse because the word is allowed to float. One speaker means secrecy, another means anonymity, while a third means legal protection against search. The result is not disagreement so much as collision.

Eric Hughes gave the clean definition this book adopts: privacy is the power to selectively reveal oneself to the world. The load-bearing word is *selectively*.

Privacy means controlling disclosure to the people who have a proper claim to the information, for the purposes that require it. A patient who speaks openly with a doctor is exercising privacy. So is a merchant who keeps supplier terms from competitors. These cases do not involve absolute concealment. They involve control over who knows what, and that control is active. Privacy is the ability to decide what information moves outward and what stays put, and if that ability disappears, privacy is gone even when no one has yet exploited the exposed data. The cost begins with the loss of control: every later decision the actor makes faces a wider range of possible adversary response.

The definition also keeps the subject clear. Privacy concerns information about persons: plans, preferences, correspondence, health, finances, relationships, political commitments, location, and the thousand other details that make a life legible to those who would exploit it. The book will later argue that information can not be property in the strict sense. But it does not follow that personal information is free for the taking.

This point dissolves the cheap force of the "nothing to hide" argument. Everyone manages disclosure, and everyone reveals different facts to different people. A person who speaks one way to his accountant and another in public is not acting suspiciously; he is acting like a person.

1.2 Three Related Terms

The next confusion comes from treating three distinct ideas as if they were interchangeable. Privacy concerns controlled disclosure; secrecy concerns concealment; anonymity concerns attribution. They overlap, but they do not collapse into each other.[3]

Privacy is the broadest of the three. It covers the ordinary management of information inside identified relationships. You may be fully known to your doctor and still have privacy if your records

remain under proper control. You may be identified to a trading partner and still preserve privacy if only the terms needed for the trade are shared.

Secrecy is narrower and harder than privacy. A secret is hidden from everyone except the few who are meant to know it, or from everyone altogether. Trade secrets, source protection, intelligence methods, and private keys all rely on secrecy, but some private matters are secret and many are not.

Anonymity concerns the link between action and actor. An anonymous donation or an unlinked payment does not tell the observer who performed the act; a pseudonymous essay does so only under a chosen identity. A person may preserve anonymity in one domain while being fully identified in another. The question is not what is known about the person in general, but whether this act can be tied to that person.

Surveillance exploits their confusion.

If privacy is collapsed into secrecy, the defender is forced into the position of explaining why he hides things.

If anonymity is collapsed into criminal evasion, the defender must explain why any unattributed action should be tolerated.

If secrecy is treated as the whole of privacy, ordinary control over personal information starts to look extreme.

The better view is direct.

Privacy is ordinary and constant.

Secrecy is a narrower case that some activities require.

Anonymity is a tool that becomes necessary when attribution itself creates danger.

Many real situations use several of these tools at once. A journalist protecting a source may need privacy over communications and secrecy about source identity. A buyer using a privacy-preserving payment system may not seek total secrecy at all. He may want only to keep merchants and banks from watching his complete transaction history.

The argument against privacy cheats when the three words slide together.

1.3 Privacy as Strategic Defense

Privacy is a moral good and a strategic defense.

John Boyd's OODA loop, Observe, Orient, Decide, Act, provides a simple way to see why.[4] An adversary who wants to control, seize, punish, censor, or preempt must first observe. Without observation the rest of the sequence stalls. A state that cannot see funds cannot freeze them on command, a platform that cannot inspect messages cannot rank or suppress them with the same precision, and a regulator who cannot map the network cannot identify the soft points where pressure will hurt most.

Observation is why the fight over privacy begins so early in the chain. It looks passive. But it is not. Observation is what lets every later intervention become targeted and cheap.

The defender's advantage lies exactly here. Surveillance requires infrastructure, storage, indexing, legal authority, software, analysts, and time. In many cases the countermeasure is far cheaper: a key pair costs almost nothing, and end-to-end encryption can run on hardware people already own. A privacy-preserving communication tool can deny easy collection at a fraction of the attacker's cost.[5]

That asymmetry matters more than the slogan that privacy is a right. Rights language is important, but costs still decide much of political reality. A state can maintain control when observation is cheap. It struggles when observation becomes expensive and unreliable.

This book will return to that asymmetry many times. It is one reason privacy tools provoke such hostility from institutions built on surveillance. They do not hide information. They raise the price of domination.[6]

1.4 The Book's Argument in Brief

This book develops one argument through several layers. Privacy belongs to the structure of action itself. It also supports economic coordination by protecting deliberation and negotiation, along with the signals on which markets depend. Surveillance is

therefore not a neutral administrative convenience. It is a method of control used by states and allied institutions to tax, redirect, or suppress. The answer is practical as well as theoretical: privacy can be defended through cryptography, anonymous communication, resistant money, selective disclosure, and decentralized protocols. Those tools allow the growth of a parallel economy that depends less and less on systems built for observation.[7]

The order of the book follows that logic. Part II develops the axioms, Part III applies economic reasoning, Part IV studies the adversary, and Parts V and VI turn to tools, implementation, institutions, and the parallel economy itself.

The book also states its limits.

It does not claim that privacy removes every danger or that technology makes politics irrelevant. Nor does it claim that parallel institutions replace the official order overnight. It claims something narrower and stronger. Privacy is defensible in theory. It is possible in practice and becomes economically potent once enough people know how to use it.

The book also marks one further scope. Privacy is not a separate property right alongside ownership of one's body and ownership of scarce external resources. What ordinary language calls "privacy" is what self-ownership of the body, property rights in physical devices and premises, and contractual confidentiality produce when respected. These three rights are common ground in the libertarian-Austrian tradition: Locke and Rothbard ground them in self-ownership and original appropriation;[8] Hoppe grounds them in the presuppositions of argumentation;[9] Kinsella sharpens them through the scarcity criterion.[10] The book's privacy argument rests on what these positions share. Chapter 6 develops the framework in detail and follows Kinsella's later treatment because it is more precise about what contract can and cannot do, while marking where Rothbard's older treatment differs and what the differences reach. The chapters in between build the case for protecting the underlying rights and for the engineering that defends them in practice.

Chapter Summary

Privacy is control over disclosure, not concealment, the power to selectively reveal oneself. That definition separates privacy from two adjacent concepts. Secrecy is narrower, and anonymity concerns attribution. The "nothing to hide" argument works only by collapsing these three into one; once they are held apart, it loses its force.

Privacy is also strategic. Observation is the first step of every targeted intervention, and raising its cost disrupts the whole sequence. Surveillance requires infrastructure, analysts, and time; a key pair costs almost nothing. That asymmetry, more than any appeal to rights, explains why privacy tools provoke such hostility from institutions built on surveillance.

What remains for later is the ground beneath these claims. Privacy as a structural feature of action waits for Chapter 3, the normative case that surveillance is unjust for Chapter 4's argumentation ethics, and the economic consequences of observed exchange for Chapters 7 and 9. This chapter fixes the vocabulary those arguments presuppose.

Notes

[1]Eric Hughes, "A Cypherpunk's Manifesto" (March 9, 1993), `https://www.activism.net/cypherpunk/manifesto.html`. The opening declaration of the cypherpunk movement, written as a short manifesto and circulated on the cypherpunks mailing list. Hughes's working definition of privacy as "the power to selectively reveal oneself to the world" supplies the operational definition this chapter adopts; the line in the epigraph names the structural claim, that privacy is a precondition of an open society in the electronic age, not a freestanding right that competes against other social goods. For the foundational mid-twentieth-century academic treatment of privacy as control over information, see Alan F. Westin, *Privacy and Freedom* (New York: Atheneum, 1967), which established the standard four-function taxonomy (solitude, intimacy, anonymity, reserve) and is the standard reference for privacy scholarship in the legal-academic tradition.

[2]The definitive scholarly treatment of the "nothing to hide" argument is Daniel J. Solove, "'I've Got Nothing to Hide' and Other Misunderstandings of Privacy," *San Diego Law Review* 44 (2007): 745, available at `https://papers.ssrn.com/abstract=998565`. Solove catalogs the rhetorical moves that let the slogan succeed (equating privacy with concealment,

treating privacy-as-individual-right in conflict with security-as-collective-good, demanding specific harms before granting privacy standing) and shows why each fails. Book-length development in Solove, *Nothing to Hide: The False Tradeoff between Privacy and Security* (Yale University Press, 2011), which surveys the surveillance-state arguments that became dominant after 2001 and explains why the "I have nothing to hide" response concedes the entire framing. For the lived-experience refutation, Aleksandr Solzhenitsyn, *The Gulag Archipelago*, trans. Thomas P. Whitney (New York: Harper & Row, 1974), documents the Soviet-era truth that under sufficient state scrutiny any individual can be found to have something to conceal; Solzhenitsyn's observation is that the "nothing to hide" defense presupposes a benign state whose behavior under pressure is not historically attested.

[3]On the conceptual taxonomy of privacy, see Westin (cited in note 1 above), whose four-function taxonomy and control-based definition are the modern starting point. Daniel J. Solove, *Understanding Privacy* (Harvard University Press, 2008), extends Westin's work and analyzes privacy through a taxonomy of sixteen distinct harms including surveillance, aggregation, identification, and disclosure; it is the standard modern scholarly treatment. Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press, 2010), develops the contextual-integrity framework that explains why disclosures appropriate in one context (health information to a doctor) become violations in another (the same information to an employer). Anita L. Allen, *Unpopular Privacy: What Must We Hide?* (Oxford University Press, 2011), treats privacy rights from a legal-philosophical angle and makes the case that some privacy claims are owed regardless of the claimant's preferences. Julie E. Cohen, "What Privacy Is For," *Harvard Law Review* 126 (2013): 1904–1933, adds that privacy shelters the emergent subjectivity of the individual from the efforts of commercial and government actors to render it fixed, transparent, and predictable. For the constitutional-law origin of the American privacy-rights tradition, see Louis D. Brandeis's dissent in *Olmstead v. United States*, 277 U.S. 438, 478 (1928), which established "the right to be let alone, the most comprehensive of rights and the right most valued by civilized men" as the standard formulation; and Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4 (1890): 193–220, the founding article of American privacy jurisprudence.

[4]John Boyd developed the OODA loop through a series of military briefings, especially "Patterns of Conflict" and "The Strategic Game of ? and ?"; his slides are archived at the Defense and the National Interest project (`https://www.projectwhitehorse.com/pdfs/boyd/patterns%20of%20conflict.pdf`). For a book-length treatment, see Frans P. B. Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd* (Routledge, 2007), which is the standard academic study of Boyd's work and traces the OODA loop from fighter-pilot tactics through grand strategy. Chet Richards, *Certain to Win: The Strategy of John Boyd, Applied to Business* (Xlibris, 2004), is the best non-military application. For the original briefing cycle compiled and edited, see Grant T. Hammond, *The Mind of War: John Boyd and American Security* (Smithsonian, 2001).

[5]The economic asymmetry between cheap encryption and expensive attack is treated in Bruce Schneier, *Applied Cryptography*, 20th anniversary ed. (Wiley, 2015), and in Niels Ferguson, Bruce Schneier, and Tadayoshi Kohno, *Cryptography Engineering: Design Principles and Practical Applications* (Wiley, 2010). For the policy consequences, see Whitfield Diffie and Susan Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption*, updated ed. (MIT Press, 2007), and Susan Landau, *Listening In: Cybersecurity in an Insecure Age* (Yale University Press, 2017). The foundational public-key paper is Whitfield Diffie and Martin Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* IT-22, no. 6 (1976): 644–654, which argued that asymmetric cryptography changes the economics of secure communication by removing the key-distribution bottleneck.

[6]The political economy of state and corporate surveillance is treated at book length in Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2019), which frames surveillance as a business model that extracts behavioral data as its primary commodity. Carissa Véliz, *Privacy Is Power: Why and How You Should Take Back Control of Your Data* (Bantam Press, 2020), makes the shorter popular case. Bruce Schneier, *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World* (Norton, 2015), surveys the technical and political infrastructure. For the historical frame that underlies the book's account of why states pursue legibility, James C. Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (Yale University Press, 1998), documents how legibility itself is the prerequisite for state intervention and how resistant populations have used opacity as a political strategy. On the cryptocurrency-specific surveillance industry, Andy Greenberg, *Tracers in the Dark: The Global Hunt for the Crime Lords of Cryptocurrency* (Doubleday, 2022), is the definitive journalistic account of blockchain forensics as practiced by the firms this book's Chapter 20 treats.

[7]Further reading on privacy as a field: Julia Angwin, *Dragnet Nation: A Quest for Privacy, Security, and Freedom in a World of Relentless Surveillance* (Times Books, 2014), documents one reporter's practical attempt to disentangle from surveillance infrastructure, and is a useful companion to Chapter 23. Finn Brunton and Helen Nissenbaum, *Obfuscation: A User's Guide for Privacy and Protest* (MIT Press, 2015), catalogs the techniques of adding noise to signals an adversary wants clean, and their ethics. Glenn Greenwald, *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State* (Metropolitan Books, 2014), remains the best narrative account of the 2013 NSA revelations. Simon Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography* (Anchor, 2000), is the most accessible general introduction to cryptographic history. Andy Greenberg, *This Machine Kills Secrets: How WikiLeakers, Cypherpunks, and Hacktivists Aim to Free the World's Information* (Dutton, 2012), traces the cypherpunk lineage that Chapter 2 develops.

[8]John Locke, *Two Treatises of Government*, ed. Peter Laslett (Cambridge: Cambridge University Press, 1988; orig. 1689), Second Treatise, Chapters V–IX, `https://www.gutenberg.org/ebooks/7370`. Locke grounds property rights in self-ownership: each person owns his body and therefore owns the labor

of his body, and by mixing that labor with unowned resources acquires rightful title. Murray N. Rothbard extends and radicalizes this framework in *The Ethics of Liberty* (New York: New York University Press, 1998; orig. 1982), https://mises.org/library/book/ethics-liberty, arguing that all rights reduce to self-ownership and original appropriation, that the state violates both, and that contractual confidentiality is a direct application of property rights over one's own person and information.

[9]Hans-Hermann Hoppe, *The Economics and Ethics of Private Property: Studies in Political Economy and Philosophy*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2006), https://mises.org/library/book/economics-and-ethics-private-property, especially "The Ultimate Justification of the Private Property Ethic." Hoppe's argumentation ethics holds that the act of arguing presupposes exclusive control over one's body and the resources one uses to argue; any attempt to deny property rights is therefore self-refuting in discourse. Chapter 4 of this book applies this framework directly to privacy: the capacity for rational discourse presupposes the ability to control what one discloses to interlocutors.

[10]Stephan Kinsella, *Against Intellectual Property* (Auburn, AL: Ludwig von Mises Institute, 2008), https://mises.org/library/book/against-intellectual-property; and Kinsella, "How We Come to Own Ourselves," *Mises Daily* (September 7, 2006), https://mises.org/mises-daily/how-we-come-own-ourselves. Kinsella's contribution is to ground property rights exclusively in the scarcity of physical resources: only scarce goods can generate real conflict over use, and therefore only scarce goods can be the object of property rights. This scarcity criterion excludes information from the category of property in the strict sense (a conclusion this book accepts) while preserving full property rights in the physical media and devices that store or transmit information, which is what the privacy argument requires.

2

TWO TRADITIONS, ONE CONCLUSION

> "In a crypto-anarchy the government is not temporarily destroyed but permanently forbidden and permanently unnecessary."
>
> Wei Dai, "b-money" (1998)[1]

This book joins two traditions that developed apart from each other and still arrived at the same place.

The Austrian tradition approached privacy through logic, starting from the fact of human action and working toward conclusions about property, exchange, calculation, and coercion. The cypherpunk tradition approached privacy through engineering, starting from the technical possibility of secure communication and building tools that let people act without asking permission. One tradition explained why privacy matters; the other showed how privacy can be defended in practice.

Their agreement is not accidental. Both were forced to confront the same problem, which is how human beings coordinate when they act on their own plans and when other people try to interrupt, tax, monitor, or control those plans. Austrians met that problem through praxeology, and cypherpunks met it through code. They reached the same answer because they were studying the same reality from different angles.

Chapters 3 through 5 develop the axiomatic core; this chapter shows why the two traditions belong in the same book and why their synthesis produces something stronger than either line alone.

2.1 The Austrian Approach: Deduction from Action

Carl Menger and Methodological Individualism

The Austrian School begins with the acting individual. Carl Menger's *Principles of Economics* broke with the historical school by refusing to treat collective abstractions as primary.[2] Only individuals act. Markets, societies, and states are the patterns their actions produce.

That commitment is methodological, not rhetorical. If social phenomena arise from the choices of individuals, analysis must begin there; prices, institutions, and monetary arrangements are all downstream of action, and whatever stability and scale they acquire, they remain the cumulative result of many separate decisions.

The same methodology applies to privacy. If action begins in the person, then knowledge relevant to action also begins there. Plans form inside minds before they appear in speech or trade, and preferences are not public facts waiting to be collected but interior judgments that actors reveal only under chosen conditions.

Menger also identified spontaneous order. Money, language, and market coordination do not need a designer standing above the process; they emerge through repeated interaction, and the point will matter later when this book turns to the parallel economy. If order can emerge without a central architect, then privacy-preserving order can emerge without state supervision as well.

Ludwig von Mises and Praxeology

Mises turned these insights into a method. Praxeology begins from a simple fact, namely that human beings act purposefully, and the claim is not statistical but built into the act of trying to deny it: anyone who argues that action is not purposeful is still using purposeful behavior to make the case.

From that starting point, Mises derived a body of economic reasoning without waiting for surveys or laboratory confirmation. Action implies choice, preference, and opportunity cost. Exchange implies expected gain, while calculation requires meaningful signals. The whole structure unfolds from the inner logic of action itself.

That structure has direct consequences for privacy. Deliberation is internal before it is external; preferences are subjective before they are observed; plans exist before they are disclosed. A world in which every choice were visible before it was made would not be a world of free action. It would be a world where every plan could be blocked or exploited by anyone strong enough to observe it.

Chapter 3 develops this argument fully. The point here is simpler: the Austrian method treats privacy as built into the conditions of action, not as an optional sentiment or bourgeois preference.

Property and Argument

Rothbard extended praxeology into political theory by linking action to self-ownership and property[3], then to aggression.[4] Hoppe later showed that some of the same conclusions are already presupposed in argumentation itself.[5] Their routes differ, but their destination overlaps.

For this book, the key point is that surveillance is not neutral observation. When a state compels disclosure, searches a house, monitors correspondence, or builds a dossier from the traces of ordinary life, it enters domains the actor did not offer for public use. It turns private judgment into an object of appropriation.

Rothbard framed this through natural rights; Hoppe framed it through the presuppositions of discourse. Chapter 4 takes the Hoppean route because it fits the book's method more tightly. The Hoppean argument begins from what a person must already presuppose when entering argument and making claims on others, not from moral intuition.

The institutional implication is already visible. A society built on action and exchange, then willing to argue about rules, sits uneasily with universal observation. A regime that treats every transaction and every message, along with the associations they

expose, as inspectable by default moves away from order by consent and toward domination.

Konkin and Counter-Economics

Samuel Edward Konkin III translated Austrian analysis into strategic practice. He asked the question that many theorists avoid: if the state is parasitic, what do free people do before it disappears?

His answer was counter-economics[6]: build outside the channels the state can easily see and tax, trade where surveillance is weak, use forms of money the state cannot freeze at will, and shift production and exchange into spaces where voluntary order can prove itself in practice.[7]

This move brought Austrian theory to the edge of cypherpunk strategy even though Konkin wrote before Bitcoin, Nostr, or modern anonymous networks existed. The logic is already there in his argument. If the state feeds on visibility, then unseen exchange weakens it; if regulation depends on chokepoints, then systems that route around chokepoints loosen its grip; and if markets can solve coordination problems on their own, then free people do not need permission to begin building the alternative.

Konkin's importance to this book lies here. He turns Austrian critique into a theory of transition. He points from logical analysis to parallel practice.

What the Austrian Tradition Contributes

The Austrian line contributes three things that the rest of the book will develop in depth.

It gives a descriptive claim: privacy belongs to the structure of action.

It gives the basis for a normative claim developed later: coercive intrusion into private domains is not morally innocent.

It gives a strategic claim: free order can emerge through voluntary coordination without central direction.

The Austrian contribution is already enough to establish privacy's place in action, property, and exchange. It is not yet enough

to defend privacy in the real world. For that, the second tradition becomes necessary.[8]

2.2 The Cypherpunk Approach: Building What Theory Requires

David Chaum and Privacy by Design

David Chaum saw the problem before most of the world understood the stakes.[9] Early networked systems were already creating the conditions for total recordkeeping. Chaum's answer was architectural: tools that would render the surveillance assumption inoperative without waiting on legislatures or courts.

Blind signatures showed that verifiable digital payment did not require the issuer to know who spent the token. That insight is larger than the primitive itself. It changed the terms of the debate. The question stopped being whether institutions could be trusted to behave well after collecting data. The question became whether the system could be designed to avoid collecting the data in the first place.

That is a core cypherpunk move. Do not beg power to limit itself. Remove the power.

The same logic will appear again and again in this book. Hope is a poor basis for privacy, so the discipline is always to remove the capability that would require hope in the first place. Instead of hoping the bank stays honest with your transaction history, use money the bank cannot control. Instead of hoping the platform resists censorship when pressure arrives, use protocols no platform owns. Instead of hoping the database stays secure after each compromise or subpoena, refuse to build the dossier at all. In every case, the cypherpunk instinct is to remove the capability at the source and not to trust the institution to decline to use it.

Software Freedom and Verifiability

The cypherpunk answer also required one further condition: users must be able to inspect the tools they trust.

Here Richard Stallman's importance is monumental.[10] Software that cannot be studied cannot be trusted in a strong sense. A closed binary may promise privacy while exfiltrating keys, logging metadata, or baking in dependencies the user never approved. Open code does not guarantee security, but closed code blocks independent verification from the start.

Cypherpunk systems ask users to trust mathematics and public scrutiny more than institutional benevolence, and that trust collapses if the code cannot be inspected and maintained outside one vendor's control, with forking available if needed. Every major privacy system in the modern stack, including Tor, Bitcoin, Signal, and Nostr clients, depends on this assumption.[11]

The link to Austrian analysis is tight. Information is not scarce, and voluntary order requires the freedom to inspect what one relies on. Hidden control inside closed software recreates dependency under a technical skin.

Timothy May and the Politics of Anonymous Commerce

Timothy May stated the political implications with unusual clarity. If cryptography lets people communicate and transfer value without showing legal identity, then the state's ordinary tools of taxation and control lose much of their force.[12]

May also saw the next problem. Anonymous business does not abolish trust; it changes how trust must be built, because reputation becomes central once legal identity stops doing the usual work.[13] That observation is one of the most important bridges in the whole manuscript: the cypherpunk project was never only about secrecy of messages, but also about the conditions for economic coordination once identity and payment move outside the state's default channels.

This book follows May all the way to that institutional question. Chapter 24 is one answer to the problem he saw.

Eric Hughes and Selective Disclosure

Eric Hughes gave the movement its cleanest ethical summary.[14] Privacy means selective disclosure. Each party to a transaction

should know only what is directly necessary for that transaction. The cypherpunk line is visible here in full.

The discipline is to ask what the exchange requires, not to collect broad identity data and promise restraint later. If the transaction needs proof of payment, then prove payment; if it needs proof of age, then prove age; if it needs proof of continuity with a known key, then prove that continuity. Anything beyond the real informational requirement is surplus extraction that the protocol should refuse to produce in the first place.

Hughes also supplied the movement's governing method: cypherpunks write code. Arguments become harder to reverse once they are embodied in working systems. A court ruling can be overturned. A legislative win can be repealed. A protocol with a real user base and a distributed set of operators is a different kind of fact.

From Cyberspace Independence to Parallel Institutions

John Perry Barlow's declaration captured the early optimism of a borderless internet.[15] Later texts grew harder and more disciplined. What remained after the optimism burned off was the durable strategic insight: freedom can be practiced in parallel through separate rails of communication, payment, and association, without waiting for a final political break.

Here the cypherpunk line meets Konkin's counter-economics. Anonymous messaging, digital cash, open protocols, and cryptographic proof function as the infrastructure of separation as well as the technical novelties they appear to be. They let people begin building markets and communities that do not ask every question through the state's categories.

Smuggler and XYZ carried this line into strategic practice. *The Second Realm* treats cypherpunk tools as the working infrastructure of parallel institutions: protected digital channels and recurring physical meeting places, with proxy merchants that let trade move across the boundary under lower exposure. In that sense it is modern cypherpunk strategy applied to ordinary life.[16]

The mature version of this strategy aims at something harder than digital escapism and does not imagine a clean migration into a separate cyber-domain. It aims at functioning institutions that coexist with the official order while depending on it less and less.

Digital Cash Precursors and the Cypherpunk Method

A secure and private digital money has been a dream of the cypherpunks from the start. Before Bitcoin, they solved parts of the puzzle one by one. Hashcash showed how computation could impose cost.[17] B-money described participant-created digital money, and Bit Gold pushed digital scarcity closer to a usable design.[18] Reusable proof-of-work[19] and related experiments filled in more of the map.[20]

Even the systems that failed reveal the cypherpunk method: build, publish, test against reality, and keep what survives.

The cypherpunk method differs from praxeology without being in conflict with it. Cypherpunks answer practical questions that theory alone cannot settle, such as which cryptographic assumptions hold up under attack, which network designs resist censorship in hostile conditions, and which monetary systems can survive adversarial pressure and real user mistakes. Those questions need code, not deduction alone.

What the Cypherpunk Tradition Contributes

The cypherpunk line contributes three things.

It shows that privacy can be built into systems as a property of architecture.

It shows that verification cannot rest on trust in opaque institutions, which forces tools to be open and inspectable.

It shows that the path to freedom runs through deployed systems, not declarations alone.

The cypherpunk contribution is already enough to explain how privacy can be defended. It is not yet enough to explain why privacy belongs to the structure of human coordination. For that, the Austrian line remains necessary.[21]

2.3 Independent Convergence: Why Both Reached the Same Place

Different Methods, Same Constraints

The Austrian and cypherpunk traditions used different methods because they confronted different parts of the same problem.

Austrians asked what must be true if action and exchange are real under conditions of calculation. Cypherpunks asked what must be built if privacy and sound money, together with free communication, are to survive hostile conditions. One line moved through deduction. The other moved through implementation.

The convergence is still striking.

Each tradition begins with the person, not the collective abstraction; rejects the idea that central oversight improves coordination as such; and discovers that coercion distorts the signals on which order depends. Both end up defending decentralization and voluntary exchange against unchecked observation.

That agreement does not require one tradition to borrow its conclusions from the other. The stronger explanation is that both had to reckon with the same constraints of action and control.

Shared Conclusions

The shared conclusions can be stated plainly.

Human beings act on private judgments before those judgments become public.

Exchange works better when parties can choose what to disclose.

Free order can emerge without a master plan.

Surveillance and coercion damage the process they claim to improve.

A tradition that starts with action is pushed toward privacy. A tradition that tries to defend freedom in code is pushed there too.

What the Synthesis Adds

Each tradition alone leaves something exposed.

Austrian theory can derive privacy from action and can show why intervention degrades coordination. By itself it cannot identify which tools resist surveillance in the present world.

Cypherpunk engineering can build communication and payment systems that work under hostile conditions, along with identity systems suited to them. By itself it can still lose its way economically, with projects centralizing and incentives drifting, and systems that look elegant in code failing because they ignore human coordination or capital structure.

The synthesis closes both gaps.

Austrian analysis locates privacy inside human action and specifies the incentives that privacy-preserving systems must respect. Cypherpunk practice shows how those systems can be built and defended against real adversaries.

That combination gives this book its shape. The chapters ahead move from axiom into economics, then through the adversary and the tool stack before they reach practice. The two traditions do not agree only in the abstract. They complete each other in sequence.

Chapter Summary

Two traditions reached the same conclusion about privacy from different starting points. The Austrian line begins with the acting individual: from Menger through Mises, Rothbard, Hoppe, and Konkin, it shows that action starts in private judgment, that voluntary order can emerge without central design, and that coercive intrusion damages the conditions of exchange. The cypherpunk line begins with technical possibility: from Chaum through Hughes and May, then into the digital cash precursors, it shows that privacy can be built into systems, that open code matters for verification, and that working tools can shift power away from institutions built on observation.

Their convergence is evidence. They used different methods because they confronted different sides of the same reality, and each contributes what the other lacks. Austrian analysis grounds privacy in action and identifies the voluntary order that makes it

possible; cypherpunk engineering shows how resistant systems can be built and maintained against real adversaries. Konkin's counter-economics is the bridge, the Austrian critique turned into a program of transition that the cypherpunks then made operational.

This chapter traces intellectual lineage. The full axiomatic derivation begins in Chapter 3, the normative case in Chapters 4 and 5, and the assessment of Bitcoin, Tor, Nostr, and the rest in Parts V and VI.

Notes

[1]Wei Dai, "b-money" (November 1998), `http://www.weidai.com/bmoney.txt`. A two-page proposal for an anonymous digital currency that Satoshi Nakamoto cited as reference [1] in the Bitcoin whitepaper. Dai's opening paragraph, quoted in this chapter's epigraph, names the fusion of cryptography with economic thought that this chapter traces: privacy infrastructure designed not as policy reform but as permanent architecture of exchange. For the rhetorical companion text that named the movement, see Eric Hughes, "A Cypherpunk's Manifesto" (1993), `https://www.activism.net/cypherpunk/manifesto.html`, which supplied the phrase "cypherpunks write code" and is quoted throughout the chapter.

[2]On Menger's method and the Austrian break with the historical school, see Carl Menger, *Principles of Economics* (1871), and Ludwig von Mises, *Human Action* (1949), especially the methodological chapters.

[3]Murray N. Rothbard, *Man, Economy, and State: A Treatise on Economic Principles*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 2009 [1962]). Rothbard's most systematic work, deriving price theory, capital theory, and the theory of production from the action axiom. Where *The Ethics of Liberty* grounds property rights in self-ownership and natural law, *Man, Economy, and State* grounds them in the logic of action and exchange, making it the core economic complement to the political theory cited in note 3.

[4]Murray N. Rothbard, *The Ethics of Liberty* (Humanities Press, 1982), develops the derivation of self-ownership and property from action and the concept of aggression as boundary-crossing.

[5]Hans-Hermann Hoppe, *The Economics and Ethics of Private Property: Studies in Political Economy and Philosophy*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2006), `https://mises.org/library/economics-and-ethics-private-property`. The definitive collection of Hoppe's argumentation-ethics derivation of property rights and its applications. Hoppe shows that anyone who argues against self-ownership already presupposes control over their own body and reasoning capacity to conduct the argument — a performative contradiction that grounds

property rights without appeal to moral intuition. The argumentation-ethics chapter (originally published 1988) is reprinted here alongside applications to money, banking, and state theory.

[6]Samuel Edward Konkin III, *An Agorist Primer* (KoPubCo, 2008), https://kopubco.com/pdf/An_Agorist_Primer_by_SEK3.pdf. Konkin's shorter and more accessible introduction to counter-economics and agorism, written as a companion to *New Libertarian Manifesto*. Where the Manifesto makes the strategic and philosophical case, *An Agorist Primer* explains the economic logic of grey and black markets as the practical terrain on which voluntary order expands at the state's expense.

[7]For Hoppe's argumentation ethics, see Hans-Hermann Hoppe, *A Theory of Socialism and Capitalism* (1989), Chapter 7. On Konkin and counter-economics, see Samuel Edward Konkin III, *New Libertarian Manifesto* (1980).

[8]Readers who want to work through the Austrian tradition in depth should begin with Ludwig von Mises, *Human Action* (1949), still the single most complete exposition of praxeology. Murray Rothbard, *Man, Economy, and State* (1962) extends and systematizes Mises. For a shorter entry, Peter Boettke, *Living Economics* (2012) and Jesús Huerta de Soto, *The Austrian School* (2008) are excellent. Carl Menger, *Principles of Economics* (1871) and *Investigations into the Method of the Social Sciences* (1883) are the original sources. For the relationship between Austrian method and cryptographic systems specifically, Eric Voskuil's *Cryptoeconomics* (online) applies praxeology directly to Bitcoin.

[9]David Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Communications of the ACM* 24, no. 2 (1981): 84–88, https://nakamotoinstitute.org/library/untraceable-electronic-mail/, introduced mix networks; Chaum, "Blind Signatures for Untraceable Payments," in *Advances in Cryptology: Proceedings of Crypto '82* (Plenum, 1983), 199–203, introduced the blind-signature primitive discussed in this section; Chaum, "Security Without Identification: Card Computers to Make Big Brother Obsolete," *Communications of the ACM* 28, no. 10 (1985): 1030–1044, https://nakamotoinstitute.org/library/security-without-identification/, generalized the design philosophy.

[10]Richard Stallman, "The GNU Manifesto" (1985), https://www.gnu.org/gnu/manifesto.html, and the Free Software Foundation, https://www.fsf.org/. The four freedoms are stated at https://www.gnu.org/philosophy/free-sw.html.

[11]Canonical project pages for the systems named in this paragraph, each of which receives extended treatment later in the book: Tor (https://www.torproject.org), the anonymous-routing network developed from David Chaum's 1981 mix-network design and deployed since 2004, treated in Chapter 20. Bitcoin (https://bitcoin.org; whitepaper at https://bitcoin.org/bitcoin.pdf), the peer-to-peer electronic cash system introduced by Satoshi Nakamoto in October 2008 and launched in January 2009, treated in Chapter 18 and Chapter 19. Signal (https://signal.org), the end-to-end encrypted messenger whose Double Ratchet protocol is the reference

design for modern secure messaging, treated in Chapter 17; the protocol specification is at `https://signal.org/docs/`. Nostr (`https://nostr.com`; protocol specifications in the NIPs repository at `https://github.com/nostr-protocol/nips`), the public-key-based decentralized social protocol introduced by fiatjaf in November 2020, treated in Chapter 21.

[12]Timothy C. May, "The Crypto Anarchist Manifesto" (1988), `https://nakamotoinstitute.org/library/crypto-anarchist-manifesto/`. For the fuller treatment of persistent digital pseudonyms, anonymous business, and reputation inside virtual communities, see May, "Crypto Anarchy and Virtual Communities" (1994), `https://nakamotoinstitute.org/library/virtual-communities/`.

[13]On the libertarian-property treatment of reputation that May's observation rests on: reputation is not a resource in the property-theoretic sense, since it consists of what other people believe about a person, not in any resource the person owns. Cypherpunk reputation systems (developed in Chapter 21) do not treat reputation as ownable; they make signed attestations and consistent behavior over time the empirical signals on which counterparties form their own assessments. For the framework, see Stephan Kinsella, "Defamation as a Type of Intellectual Property," in Jörg Guido Hülsmann and Stephan Kinsella, eds., *A Life in Liberty: Liber Amicorum in Honor of Hans-Hermann Hoppe* (Mises Institute, 2024), and the treatment in Chapter 6, §6.5 of this book.

[14]Eric Hughes, "A Cypherpunk's Manifesto" (March 9, 1993); see note 1 above for the full citation and URL.

[15]John Perry Barlow, "A Declaration of the Independence of Cyberspace" (February 8, 1996), `https://www.eff.org/cyberspace-independence`. Written at the World Economic Forum in Davos two days after President Clinton signed the Telecommunications Reform Act, the declaration asserts that cyberspace constitutes a separate social space beyond state sovereignty: "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone." The declaration's optimism about voluntary self-governance and its rejection of territorial jurisdiction were not sustained as predictions; as a statement of the movement's early self-understanding and aspirations, the text has no equal.

[16]For a later strategic text that carries the cypherpunk line past anonymous messaging and digital cash into recurring parallel institutions, see Smuggler and XYZ, *The Second Realm: Book on Strategy* (Liberty Under Attack Publications, 2018). For the earlier autonomy language it partly inherits, see Hakim Bey, *T.A.Z.: The Temporary Autonomous Zone, Ontological Anarchy, Poetic Terrorism* (Autonomedia, 1991). The value of these works for this chapter is genealogical and strategic, not axiomatic.

[17]Adam Back, "Hashcash — A Denial of Service Counter-Measure," technical report (August 2002), `http://www.hashcash.org/papers/hashcash.pdf`. Back's proof-of-work system was originally designed to combat email spam by

requiring senders to expend CPU time. Satoshi Nakamoto cited it as reference [6] in the Bitcoin whitepaper, and Hashcash's SHA-1-based cost function directly informed Bitcoin's mining mechanism. Back later became CEO of Blockstream, a Bitcoin infrastructure company (`https://blockstream.com`).

[18]Nick Szabo, "Bit Gold" (2005), `https://nakamotoinstitute.org/library/bit-gold/`. Szabo's design for a decentralized digital scarcity system using proof-of-work chains and public ledgers anticipated Bitcoin's architecture in several respects, including the chaining of work proofs and the need for a distributed title registry. Szabo also developed the concept of smart contracts; see "Smart Contracts" (1994), `https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart.contracts.html`, and "The Idea of Smart Contracts" (1997), `https://nakamotoinstitute.org/library/the-idea-of-smart-contracts/`.

[19]Hal Finney, "RPOW — Reusable Proofs of Work" (August 2004), `https://nakamotoinstitute.org/library/rpow/`. Finney's system allowed a hashcash token to be exchanged for a signed token of equivalent value, solving the double-spending problem for proof-of-work tokens through a trusted server holding its code open to remote attestation. Finney was the first person other than Satoshi Nakamoto to run the Bitcoin software and received the first Bitcoin transaction (10 BTC in block 170, January 12, 2009). He was also a lead developer on PGP 2.0 at PGP Corporation and one of the most technically accomplished contributors to the cypherpunk mailing list.

[20]For the path from Hashcash and B-money to Bit Gold and Bitcoin, see Adam Back, "Hashcash: A Denial of Service Counter-Measure" (2002); Wei Dai, "B-money" (1998); Hal Finney, "RPOW — Reusable Proofs of Work" (2004); and Nick Szabo, "Bit Gold" (2005).

[21]On the cypherpunk tradition and its history, Steven Levy, *Crypto: How the Code Rebels Beat the Government, Saving Privacy in the Digital Age* (2001) is the standard narrative history. Andy Greenberg, *This Machine Kills Secrets* (2012) and *Tracers in the Dark* (2022) cover the later phases. Simon Singh, *The Code Book* (1999) provides the pre-digital context. Primary sources include May's "Crypto Anarchist Manifesto" (1988); Hughes's "A Cypherpunk's Manifesto" (1993); Hal Finney's "Protecting Privacy with Electronic Cash" (1993), `https://nakamotoinstitute.org/library/protecting-privacy-with-electronic-cash/`; Phil Zimmermann's "Why I Wrote PGP" (1991); and the Cypherpunks mailing list archives.

PART II

AXIOMS

3

THE ACTION AXIOM: PRIVACY AS STRUCTURAL FEATURE

> "Human action is purposeful behavior."
>
> Ludwig von Mises, *Human Action: A Treatise on Economics* (1949)[1]

The Action Axiom, formulated by Ludwig von Mises, states that human action is purposeful behavior. Mises calls the claim self-evident and a priori. Those phrases rightly trigger skepticism in readers trained to demand empirical tests for every assertion, and that skepticism deserves an answer.

Some claims are built into the concepts they use. "If A is taller than B and B is taller than C, then A is taller than C". That is true once you understand what "taller than" means; no field data is required to check it. The Action Axiom makes a claim of that kind about purposeful behavior: any attempt to argue against it must itself be purposeful behavior directed toward convincing others, so the denial performs the thing it tries to refute. This is not rhetorical trickery. Similar formulations appear across independent traditions: Tadeusz Kotarbiński's Polish praxiology (from which Mises drew the term), Confucian ethics on purposeful self-cultivation, and the twentieth-century logic of performative contradiction converge on a claim of the same shape.[2] When several traditions arrive at the same foundation by different paths, that foundation is more likely

identifying something real about action than repeating a single school's dogma.

From this axiom, we derive that privacy exists as a structural feature of human action. Deliberation is internal, preferences are subjective, and information asymmetry between actor and observer is inherent to the structure of purposeful behavior.[3] These are descriptive facts about how action works. Whether privacy should be protected is a normative question addressed in Chapter 4.

The present claim is descriptive: privacy belongs to the structure of action as human beings deliberate and choose before disclosing outward. The next chapter asks the separate normative question: if privacy is built into action, what may other people justly do to it?

3.1 The Action Axiom

The Self-Evident Starting Point

Ludwig von Mises identified the foundation of all economic analysis: human action is purposeful behavior.[4] To act is to employ means toward ends according to ideas about causal relationships. Action is not mere motion but directed effort aimed at changing circumstances from a less satisfactory state to a more satisfactory one.

This statement is self-evident in a precise sense: its denial refutes itself. To argue that action is not purposeful, one must purposefully construct an argument, purposefully select evidence, purposefully direct mental effort toward persuasion. The attempt to deny purposeful behavior is itself purposeful behavior, confirming what it seeks to deny. The pattern is not rhetorical cleverness but logical necessity: the denial is performatively self-refuting.

The Action Axiom is therefore a priori: known prior to and independent of particular experience. We do not discover it through observation but recognize it through reflection on what action entails. Any experience we could have is itself action and therefore presupposes what the axiom states.

What the Axiom Asserts

The Action Axiom asserts several things at once. Action aims at goals; the actor envisions a preferred future state and directs behavior toward achieving it. Action without purpose is not action but reflex or accident. Action also employs means toward ends: the actor perceives alternative pathways toward desired outcomes and selects among them. This selection presupposes evaluation, judging which means suit which ends.

Action involves choosing among alternatives, and to act is to give up some possibilities in favor of others; the chosen course excludes the unchosen. Finally, action is conscious behavior, and the actor is aware of what they are doing and why. Unconscious behavior, however complex, is not action in the praxeological sense.

What the Axiom Does Not Assert

Equally important is what the Action Axiom does not assert. The axiom does not claim that actions are rational in any substantive sense; it says action is purposeful, not that purposes are wise or means are effective. An actor may pursue foolish goals with inappropriate methods, and this is still action. Nor does the axiom assert that actors possess complete information. Actors act under uncertainty with incomplete knowledge; the axiom describes the structure of action, not its success.

The axiom does not assert that action is morally evaluable. It is descriptive, saying nothing about whether particular actions are good, right, or permissible. Ethical evaluation requires additional premises. Finally, the axiom does not assert that actions should be free from interference, for this would be a normative claim. The axiom describes how action works; it does not prescribe how action should be treated.[5]

3.2 Internal Deliberation and Subjective Valuation

Deliberation Occurs in the Mind

Action requires choice among alternatives. Choice requires deliberation: weighing options, considering consequences, evaluating trade-offs. Where does this deliberation occur?

It occurs in the mind of the actor.[3] This is not a contingent fact about how humans happen to work but inherent to what deliberation means. To deliberate is to engage in internal mental processes: imagining alternatives, projecting outcomes, comparing evaluations. These processes are intrinsically internal; they occur within the deliberating subject.

An external observer cannot access another's deliberation directly.[6] They can observe behavior, record statements, measure physiological responses. But the actual mental process, the weighing and evaluating that constitutes deliberation, remains internal to the deliberator. Deliberation, by its nature, is internal.

This claim about inherent inaccessibility deserves scrutiny. One might object that current inaccessibility is only a function of current technology, not an essential feature of deliberation. Future neuroscience might develop methods to "read" deliberative processes directly from brain states.[7] If so, the privacy of deliberation would be empirically contingent, not structurally guaranteed.

The move from phenomenological observation, deliberation currently appears internal to us, to structural claim, deliberation is inherently internal, is contestable. What we experience as private deliberation may someday become readable through sufficiently advanced brain imaging. This would not refute the action axiom itself, which concerns purposeful behavior, not its observability. But it would qualify the claim that information asymmetry is permanently structural and show it to be technologically contingent instead.

For present purposes, the relevant point is that deliberation is currently and for the foreseeable future internal and inaccessible. The privacy implications developed in this chapter hold given actual human capacities. Whether future technology could change this is an empirical question that does not affect current analysis.

Subjective Valuation

Action aims at goals the actor values. Value, however, is not an objective property of things but a relation between an evaluating subject and the object evaluated. The same object may be valued differently by different actors, or by the same actor at different times. Value exists only in the act of valuing.

Menger identified this insight, and Mises later developed it further: value is subjective.[8] It originates in the evaluating mind, not in the evaluated object. No "objective value" exists independent of someone's valuation. Prices emerge from the interaction of subjective valuations; they do not measure pre-existing objective values.

For privacy, this has immediate implications. An actor's valuations exist in their mind. No external observer can access another's value rankings directly. They can infer preferences from observed choices, but the underlying subjective experience of valuing remains internal and private.

Ordinal Preference Rankings

Preferences are ordinal, not cardinal.[9] Actors rank alternatives as more or less preferred, not as having measurable quantities of utility. An actor prefers A to B to C; they do not "have 50 utils from A, 30 from B, 20 from C."

This ordinal structure means preferences cannot be aggregated across individuals. No method exists to add your preference ranking to mine to produce a collective ranking. Each person's preference structure is their own, incommensurable with others'.

The privacy implication is direct: preference rankings exist only in individual minds. No external process can access or aggregate individual preferences without losing what preferences are. Collective decisions that claim to represent "social preferences" are metaphorical at best.

3.3 Information Asymmetry and Control

Structural Information Asymmetry

From the preceding analysis, a structural fact emerges: actors necessarily possess information that observers lack.[10]

The actor knows preferences and plans, along with the evaluations that rank them; an observer can only infer these from external evidence. Deliberation is experienced directly by the actor, while observers have access only to its behavioral outputs. This asymmetry is structural under ordinary human conditions: purposes and valuations begin internally, while observers reach them through inference, disclosure, or coercion.

Privacy in its most basic sense is information asymmetry between actor and observer that is built into the structure of action. Privacy exists as a descriptive fact before any normative claim is made about whether it should be protected or violated.

Control Over Disclosure

The actor, by virtue of having internal states, faces choices about disclosure: they can reveal their preferences through action or statement, they can conceal their plans by refraining from communication, and they have, in the relevant sense, control over what information about their internal states reaches others.

Such control is not absolute: others can infer preferences from observed behavior, and coercion may compel disclosure. But the baseline condition is that internal states are internal: accessible to the actor, inaccessible to others except through the actor's disclosure or others' inference.

Selective disclosure is what Hughes meant: the power to choose what to reveal and to whom.[11] The Action Axiom establishes that this capacity is built into action itself. Whether it should be protected or overridden is a separate question.

Surveillance as Externally Imposed Transparency

Surveillance attempts to overcome information asymmetry by making the actor's internal states accessible to observers. Recording behavior and monitoring communications aim to reduce the asymmetry between what actors know about themselves and what observers know about them.

The Action Axiom does not say surveillance is wrong; it says that surveillance attempts to overcome a structural feature of action. Whether such attempts succeed or should be permitted involves empirical and normative questions beyond the axiom itself.

What the axiom establishes is that the asymmetry being overcome is inherent to action. Surveillance attempts to overcome an inherent property of purposeful behavior.

Chapter Summary

Human action is purposeful behavior, and the claim cannot be coherently denied because denying it is itself purposeful behavior. From that axiom several descriptive facts follow. Deliberation is internal, occurring in the actor's mind before any outward sign. Preferences are subjective, existing only in individual acts of valuing that no external observer can reach directly. Information asymmetry between actor and observer is therefore structural, not incidental. Privacy, on this account, is a fact about how action works before it is anything else.

The chapter is descriptive and stops there. It does not claim that privacy is necessary for action to occur, people act under surveillance constantly, or that violating it is wrong, or that actors own their bodies and thoughts as property. Those are normative claims, and the Action Axiom alone cannot carry them. The Non-Aggression Principle, property rights, and the moral case against surveillance all require the argumentative foundation Chapter 4 develops from Hoppe's argumentation ethics.[12]

Notes

[1]Ludwig von Mises, *Human Action: A Treatise on Economics*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 1998), 11.

[2]Tadeusz Kotarbiński developed praxiology (spelled with an "i") independently in Poland in the 1920s–30s; see his *Praxiology: An Introduction to the Sciences of Efficient Action*, trans. Olgierd Wojtasiewicz (Oxford: Pergamon Press, 1965). Mises acknowledged Kotarbiński as the source of the term, though Mises' praxeology reaches deeper into the structure of action than Kotarbiński's narrower theory of efficient performance. For the Confucian parallel, the *Daxue* (Great Learning) and *Zhongyong* (Doctrine of the Mean) treat purposeful self-cultivation as a premise of human conduct, and the Neo-Confucian tradition following Zhu Xi developed "investigation of things" (*gewu* 格物) and "extension of knowledge" (*zhizhi* 致知) as structurally purposeful activities. For the modern logic of performative contradiction as a mode of argument, see Karl-Otto Apel, *Transformation der Philosophie* (Frankfurt: Suhrkamp, 1973), and the literature on transcendental pragmatics that follows from it.

[3]F. A. Hayek, "The Use of Knowledge in Society," *American Economic Review* 35, no. 4 (September 1945): 519–530, `https://www.econlib.org/library/Essays/hykKnw.html`. Hayek's most cited essay argues that the knowledge relevant to economic coordination is dispersed across millions of individuals and cannot be aggregated by any central planner without loss. The essay establishes the epistemic case for markets as discovery processes and grounds the claim that information asymmetry between actors and observers is not a technical deficiency but a permanent structural feature of social reality. The connection to this chapter's argument is direct: what Hayek identifies as the knowledge problem at the social level — no observer can possess what individual actors know about their own preferences, plans, and local circumstances — is the same asymmetry the Action Axiom identifies at the individual level.

[4]Mises, *Human Action*, 11–29. The opening chapters develop the action axiom and its immediate implications. See also Murray N. Rothbard, *Man, Economy, and State*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 2009), 1–77, for an alternative presentation of the same foundations.

[5]On the distinction between descriptive and normative claims in Austrian economics, see Hans-Hermann Hoppe, *Economic Science and the Austrian Method* (Auburn, AL: Ludwig von Mises Institute, 1995), particularly the discussion of praxeology as value-free science. Mises insisted that economics describes what is, not what ought to be; ethical evaluation requires additional premises.

[6]Gilbert Ryle, *The Concept of Mind* (London: Hutchinson, 1949), especially chapter 2, "Knowing How and Knowing That." Ryle's distinction between propositional knowledge (knowing that) and practical knowledge (knowing how) maps onto the deliberation/behavior gap: an observer can know that an actor performed an action while remaining entirely ignorant of the internal deliberative process — the weighing of ends and means — that constituted the

action as purposeful. Ryle's critique of what he called the "ghost in the machine" also bears on the neuroscience objection discussed in §3.2: behavioral observation reaches outputs, not the evaluative structure that produced them.

[7]The neuroscience of mental-state decoding is an active research area. For representative work on "brain reading" via fMRI and the limits of inferring intentional content from neural signals, see John-Dylan Haynes, "A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives," *Neuron* 87, no. 2 (2015): 257–270; and Poldrack et al., "Inferring Mental States from Neuroimaging Data: From Reverse Inference to Large-Scale Decoding," *Neuron* 72, no. 5 (2011): 692–700. The philosophical limits of moving from neural correlates to intentional content — the gap between brain states and the *meaning* of deliberation — are discussed in Ned Block, "On a Confusion about a Function of Consciousness," *Behavioral and Brain Sciences* 18, no. 2 (1995): 227–247. Even full neural readout would face the frame problem: identifying the intentional object of a deliberative state requires interpreting neural patterns against a background of beliefs and goals that is itself not directly observable.

[8]Carl Menger, *Principles of Economics*, trans. James Dingwall and Bert F. Hoselitz (Auburn, AL: Ludwig von Mises Institute, 2007 [1871]), 114–174. Menger's subjective theory of value revolutionized economics by locating value in the evaluating mind, not in objective properties of goods.

[9]Murray N. Rothbard, *Man, Economy, and State*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 2009 [1962]), 1–30. Rothbard's opening chapter derives the ordinal structure of preference rankings from the action axiom with greater formal precision than Mises's original treatment. The key argument: since action consists in choosing among ranked alternatives, preferences are revealed through choice, not through introspection or survey; cardinal measures of utility ("50 utils") presuppose interpersonal comparisons that have no foundation in the logic of action. See also Murray N. Rothbard, "Toward a Reconstruction of Utility and Welfare Economics," in *On Freedom and Free Enterprise*, ed. Mary Sennholz (Van Nostrand, 1956), for the full critique of cardinal and welfare economics.

[10]Israel M. Kirzner, *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973), especially chapter 2. Kirzner's entrepreneurship theory focuses on the alertness that allows actors to perceive profit opportunities that others have not yet noticed — a form of knowledge that is pre-discursive, not articulable as a set of propositions, and therefore structurally inaccessible to any external observer until it is acted upon. The implication for information asymmetry: the most economically consequential knowledge actors hold is precisely the knowledge they cannot fully report even if they try, because it consists in perceiving what others have missed. See also Kirzner, *The Meaning of Market Process* (London: Routledge, 1992), for the connection between subjective knowledge, entrepreneurial discovery, and the limits of central coordination.

[11]Eric Hughes, "A Cypherpunk's Manifesto" (March 9, 1993), `https://www.activism.net/cypherpunk/manifesto.html`. The "selectively reveal" formulation is the working definition this book adopts; see Chapter 1, note 1, for the fuller citation context.

[12]The argumentation-ethics derivation of self-ownership and property rights is set out in Hans-Hermann Hoppe, *A Theory of Socialism and Capitalism: Economics, Politics, and Ethics* (Boston: Kluwer Academic Publishers, 1989), chapter 7, and developed further in Hoppe, *The Economics and Ethics of Private Property* (Auburn, AL: Ludwig von Mises Institute, 2006); Chapter 4 of this book applies that derivation to privacy. For the wider praxeology literature, the primary Austrian sources in order of increasing difficulty are Ludwig von Mises, *Human Action: A Treatise on Economics*, Scholar's Edition (Mises Institute, 1998; original 1949), the full treatment, with chapters I–VII covering the axiom and its derivations; Murray N. Rothbard, *Man, Economy, and State*, Scholar's Edition (Mises Institute, 2009; original 1962), part 1, which presents a more systematic derivation from the axiom to full economic theory; and Hans-Hermann Hoppe, *Economic Science and the Austrian Method* (Mises Institute, 1995), `https://mises.org/library/economic-science-and-austrian-method`, the shortest rigorous defense of praxeology's epistemological status against empiricist and Popperian critics. For shorter entry points, see Peter Boettke, *Living Economics: Yesterday, Today, and Tomorrow* (Independent Institute, 2012), and Jesús Huerta de Soto, *The Austrian School: Market Order and Entrepreneurial Creativity* (Edward Elgar, 2008). On a priori knowledge and its role in action theory, Barry Smith, "In Defense of Extreme (Fallibilistic) Apriorism," *Journal of Libertarian Studies* 12, no. 1 (1996): 179–192, engages the Popperian objection. Non-Austrian parallels worth reading alongside Mises include Ludwig Wittgenstein, *Philosophical Investigations* (1953), on purposive behavior and rule-following, and G. E. M. Anscombe, *Intention* (Basil Blackwell, 1957), on the philosophy of action. For the critical literature, Walter Block, Frank Arbas, and Jay Chang have collected major objections and replies in the *Journal of Libertarian Studies* and *Quarterly Journal of Austrian Economics* over several decades; Jörg Guido Hülsmann, ed., *The Last Knight of Liberalism: An Intellectual Biography of Ludwig von Mises* (Mises Institute, 2007), situates the axiomatic method historically.

4

THE ARGUMENTATION AXIOM AND SELF-OWNERSHIP

> "Any proposition must have a proposer, and the proposer's right to make his proposal must be presupposed."
>
> Hans-Hermann Hoppe, *A Theory of Socialism and Capitalism* (1988)[1]

This chapter is normative. It presents an argument for ethical conclusions.

Chapter 3 established descriptive facts: privacy is a structural feature of human action. But description alone does not yield prescription. That action has a certain structure does not, by itself, tell us what we should do about it.

This chapter develops the normative case. Hans-Hermann Hoppe argues that engaging in argumentation presupposes certain conditions, and that denying these conditions while arguing creates performative contradiction. From this, Hoppe derives self-ownership, property rights, and the Non-Aggression Principle.[2]

This argument has been debated for nearly four decades. Philosophers have raised objections, and Hoppe and his defenders have addressed them. This chapter presents the argument and examines the major objections it has faced. The is-ought problem, which has vexed philosophy since the scholastics, receives a sophisticated

solution: the argument does not derive ought from is but shows that certain normative claims cannot be coherently denied in rational discourse.

The chapter turns on three distinct questions: what argumentation presupposes while it is happening, why denying those presuppositions defeats itself, and how far the same logic reaches once the argument moves from control over body and mind to control over external resources. The third question is where most readers first feel the argument's pressure.

4.1 The Argument: Performative Contradiction

Origins in Discourse Ethics

Hoppe's argumentation ethics draws on the work of Jürgen Habermas and Karl-Otto Apel, who developed discourse ethics in the tradition of transcendental pragmatics.[3] Both sought to identify presuppositions of rational discourse, conditions that must be met for meaningful argumentation to occur.

Habermas and Apel argued that certain norms are implicit in the act of argumentation itself. If these norms are necessary for argumentation, then denying them while arguing creates what they called performative contradiction: the denial is undermined by the act of denying.

Hoppe accepted this methodological approach but rejected Habermas and Apel's conclusions about which norms discourse presupposes. Where they derived social-democratic policies, Hoppe derived self-ownership, private property, and libertarian ethics. The method is shared; the conclusions differ.

The Structure of the Argument

Hoppe's argument proceeds as follows:

Premise 1: We are engaged in argumentation. The argument applies only within this context; if we are not arguing, it does not apply.

Premise 2: Argumentation is a specific form of action with specific presuppositions. Not all action is argumentation, but all argumentation is action.

Premise 3: To argue, one must have exclusive control over one's body. You cannot argue without using your body to produce sounds, gestures, or writing. If you lacked control over your body during argumentation, you could not formulate, express, or defend your position.

Premise 4: To argue, one must have exclusive control over one's mind. Argumentation requires formulating thoughts and evaluating evidence against competing claims. If your mental processes were under another's control, you could not engage in actual argumentation.

Conclusion 1: Anyone engaged in argumentation implicitly presupposes exclusive control over body and mind. Self-ownership means precisely this: exclusive control over one's own person.

Premise 5: To deny self-ownership while arguing is to engage in performative contradiction. The denier must use exclusive control over their body and mind to formulate and express the denial. The act of denying presupposes what the denial rejects.

Conclusion 2: Self-ownership cannot be coherently denied in argumentation. Any attempt to deny it confirms it.

Extension to Property Rights

Self-ownership covers body and mind during discourse itself. The harder question is whether argumentation also presupposes rules for scarce resources outside the body, because discourse happens somewhere and uses something. Hoppe extends the argument from self-ownership to external property:

Premise 6: Argumentation occurs in time and space. Arguers occupy positions and use resources for the duration of discourse.

Premise 7: To engage in argumentation, one must have access to physical resources, at minimum a standing point from which to argue.

Premise 8: If resources could be taken from arguers during discourse, argumentation could not proceed reliably.

Premise 9: The principle that best protects argumentation is original appropriation: the first user of an unowned resource establishes property rights through mixing labor with the resource.[4]

Conclusion 3: Property rights are presupposed by argumentation. Denying property rights while benefiting from them to argue creates performative contradiction.

The Non-Aggression Principle

From self-ownership and property rights, the Non-Aggression Principle[5] follows:

If individuals own themselves and their legitimately acquired property, then uninvited interference with person or property violates ownership rights. This interference is aggression. The NAP holds that initiating such aggression is illegitimate; force is justified only in response to prior aggression.

The NAP thus rests on Hoppe's argumentation ethics, not on the Action Axiom alone. Chapter 3 could not derive the NAP because Chapter 3 was purely descriptive. The normative foundation comes here.

4.2 Objections and Responses

Hoppe's argument has faced substantial criticism. The strongest objections and their responses are examined in turn.

The Use-Ownership Gap (Murphy and Callahan)

Robert Murphy and Gene Callahan raised what many consider the most serious objection: Hoppe's argument establishes at most that arguers have use of their bodies during discourse, not that they own their bodies.[6]

The objection: "Someone can deny the libertarian ethic, and yet concede to his opponents the use of their bodies for debate. There is nothing contradictory about this."

A socialist, for instance, could say: "I grant you the use of your body for this argument. But after we finish arguing, property

arrangements will be determined collectively." This position grants temporary use without conceding permanent ownership.

Murphy and Callahan further argue that even if self-ownership is established, Hoppe's argument applies only to parts of the body used in argumentation: "At best, Hoppe has proven that it would be contradictory to argue that someone does not rightfully own his mouth, ears, eyes, heart, brain, and any other bodily parts essential for engaging in debate."

Several scholars have addressed this objection.[7] Walter Block argues that the distinction between use and ownership is artificial in this context; to have exclusive use sufficient for argumentation is to have the essential content of ownership. The socialist who says "I grant you use of your body for this argument, but afterward property arrangements will be determined collectively" has not escaped the contradiction. For the proposal itself presupposes that his body and mind are his to use in making it, that his words are his, that his position in the argument is his to defend. The "temporary permission" framing smuggles in ownership under another name.

Frank van Dun clarifies what happens when someone refuses to acknowledge these presuppositions: they place themselves outside the community of rational discourse.[8] Such a person has not "refuted" argumentation ethics but rather declined to engage in argumentation at all. They become what van Dun terms an "outlaw" in the original sense: one who has placed themselves outside the framework that makes reasoned dispute resolution possible. The argument does not claim such persons cannot exist; it establishes that they cannot coherently claim the protections they deny to others.

Hatim Kheir offers a further reformulation that strengthens the argument.[9] When parties choose arbitration to resolve disputes, they implicitly accept a framework that extends beyond immediate possession. Arbitration requires the arbitrator to decide based on objective facts and objective principles, not personal interest. The act of submitting a dispute to a neutral third party presupposes that claims can be justified through intersubjectively verifiable standards. Kheir argues that these standards necessarily include first-user acquisition and persistent ownership: if property claims vanished whenever

possession lapsed, arbitration of most real disputes would be impossible. The structure of arbitration thus presupposes ownership, not mere use.

Far from "shifting ground," Kheir's argument shows the robustness of argumentation ethics: whether one begins with argumentation in general or arbitration specifically, the same conclusions follow. The logic of rational discourse, in any of its forms, presupposes the property norms that make such discourse possible.

The Partial Application Objection

Related to the above: even if Hoppe's argument establishes some self-ownership, it applies only during argumentation. What about when people are not arguing?

A totalitarian could argue: "During this debate, you have self-ownership. Once we stop debating, different rules apply." Hoppe's argument, being about argumentation, seems to have nothing to say about non-argumentative contexts.

The response is that principles discovered through argumentation must be universalizable to function as principles at all.[10] Argumentation is the activity of providing reasons for assertions, seeking mutual understanding through rational exchange. When one proposes a norm to another person as binding on both, that norm must be universalizable to be acceptable; a particularistic norm ("I may do X to you, but you may not do X to me") provides no reason the other party could accept.

The universalizability requirement is not an arbitrary assumption but a constitutive feature of argumentation. One who offers only particularistic claims is not arguing but making assertions of power. To say "self-ownership applies during argumentation but not after" is to propose a particularistic norm: the speaker grants himself the right to suspend others' self-ownership when convenient while presumably retaining his own. This provides no reason the other party could accept and thus fails to qualify as argumentation at all.

The distinction between argumentation and mere assertion is what makes rational discourse possible. If one abandons univer-

salizability, any norm whatsoever can be asserted by inventing a particularistic exception. Without universalizability, reasoned discourse about norms collapses into assertion and counter-assertion, which is precisely what argumentation exists to transcend. The totalitarian who claims "different rules apply after the debate" has not offered an argument but a declaration of intended force.

Conflating Control and Ownership

A distinct objection: Hoppe moves from the descriptive fact that arguers control their bodies to the normative claim that they ought to have exclusive control. This conflates is and ought.

The objection: "Just as someone has the ability to control one's self, that does not give rise to why another ought to refrain from physically interfering with that control."

This objection misunderstands the structure of the argument. Hoppe does not argue: "You control your body, therefore you own it." He argues: "You cannot coherently deny self-ownership while arguing, because the denial presupposes what it denies." The ought does not enter through a derivation from is but through the requirements of non-contradictory discourse.

Consider: if someone argues "you have no right to control your body," they must use their own body to make the argument. They must presuppose their right to formulate thoughts, move their vocal cords, gesture, or type. They presuppose that their argument is theirs to make. The performative contradiction is not that control exists but that the denier must exercise the rights they deny in order to deny them.

Stephan Kinsella offers a complementary defense through the principle of estoppel.[11] An aggressor who objects to defensive force must claim that force is impermissible. But by committing aggression, he has shown through his actions that force is permissible. He is therefore estopped from objecting: to object, he would have to contradict the principle implicit in his own action. This is not deriving ought from is but showing that certain positions cannot be coherently maintained.

The Order-of-Derivation Objection

A separate and serious objection holds that the Hoppean argument has the order of derivation reversed. The Lockean-Rothbardian tradition begins with property: self-ownership is the first fact, original appropriation extends it to external resources, and the Non-Aggression Principle is the corollary that aggression against property is impermissible. On that order, property is foundational and prior; the NAP depends on a definition of property that is established first. Hoppe's argument seems to invert this: argumentation is foundational, self-ownership is derived from what argumentation presupposes, and property in external resources is derived in turn. The objection is that defining property by reference to argumentation's presuppositions is the same circularity in different clothes: the NAP still ends up depending on property, but property now depends on argumentation, which is itself a form of action that presupposes ownership of one's body. The chain has to start somewhere, and starting it at argumentation is no less arbitrary than starting it at property directly.

The honest answer is that the two orders of derivation reach the same conclusions through different routes. The Lockean-Rothbardian order takes self-ownership as a primitive moral fact and reasons outward; the Hoppean order takes argumentation as the methodologically privileged context (because every dispute about norms occurs there) and reasons that argumentation's presuppositions cannot be coherently denied within argumentation. Neither order is logically forced; both are methodological choices about where to start.

The Hoppean order has one advantage and one limitation. The advantage is that it answers Hume's is-ought challenge directly: it does not derive ought from is, it shows that certain normative claims cannot be coherently asserted in argumentation, which is a different move and one the Lockean order does not make as cleanly. The limitation is that it offers no purchase against an interlocutor who refuses to argue at all. The Lockean order has the inverse profile: it offers purchase against the silent aggressor (he has violated property, full stop) but does not directly answer Hume.

This book follows the Hoppean order because the privacy argument benefits from it. Cognitive privacy and expressive control are presuppositions of argumentation, and grounding their protection in what argumentation requires is more direct than grounding it in property derived from a separate self-ownership axiom. Either order produces the same operational conclusions: self-ownership of body and mind, original appropriation of external resources, the NAP, and the protections against coerced disclosure that this book defends. A reader who prefers the Lockean order will reach the same destination by a different path, and the chapters that follow do not require choosing between them.

Objections to the Property Extension

Even granting self-ownership, the extension to external property is separately contested. Why must arguers have private property in external resources? Why not common ownership with rules for access?

Hoppe argues that common ownership regimes cannot be universalized without contradiction. If everyone has equal access to all resources, conflicts over use are inevitable. Two people cannot occupy the same space or use the same tool simultaneously. Some resolution mechanism is needed.

The question is: what resolution mechanism can be justified through argumentation? Any proposed mechanism must be statable as a universalizable principle. "First appropriation establishes rights" is such a principle: it applies equally to all, provides clear conflict resolution, and does not presuppose prior property claims. "The collective decides" is not: it presupposes that someone has the right to speak for "the collective," that boundaries of the collective are defined, and that some mechanism exists for collective decision. Each of these presuppositions requires prior property norms to resolve.

To argue for common ownership, one must occupy a position from which to argue. One must have access to resources (a place to stand, air to breathe, a medium of communication) that others cannot simultaneously use in the same way. The arguer for common

ownership has already appropriated the resources necessary for making the argument. To then deny that appropriation establishes rights is to deny the legitimacy of the act by which the denial is made.

Left-libertarian alternatives that accept self-ownership but reject strong property rights face a further difficulty: they must explain how self-ownership can be exercised without property in external resources. To act, one must use space and materials. If these are subject to collective veto, self-ownership becomes nominal, not effective.

4.3 The Is-Ought Question

The Problem

David Hume observed that many arguments illegitimately move from statements about what is to conclusions about what ought to be.[12] Descriptive premises cannot, by themselves, yield normative conclusions: the is-ought gap.

Does Hoppe's argument bridge this gap, or does it commit Hume's fallacy?

Hoppe's Solution

Hoppe's argument does not derive ought from is in the manner Hume criticized. The structure is:[13]

1. Argumentation presupposes property in one's body and homesteading. (Descriptive claim about presuppositions)
2. Therefore, no deviation from this ethic can be argumentatively justified. (Conclusion about what can be justified)

The argument does not say: "Things are this way, therefore they ought to be this way." It says: "Anyone who enters the domain of reasoned discourse has already, by that act, presupposed the norms that make discourse possible." The ought does not come from is but from the requirements of non-contradictory rational engagement.

Murray Rothbard recognized the significance of this move, stating that Hoppe had "transcended the famous is/ought, fact/value dichotomy."[14]

Addressing Remaining Objections

Critics argue that even if certain claims cannot be denied without contradiction, this establishes only a constraint on what can be argued, not what is true. That I cannot coherently deny X while arguing does not prove X is true, only that I cannot coherently deny it.

This objection misunderstands the domain of ethical claims. Ethics concerns how rational agents ought to interact. If a proposed ethical norm cannot be coherently stated without contradiction, that is a demonstration that the norm fails as a norm, not an inconvenience for the proposer alone. A "norm" that cannot be consistently advocated is not a norm anyone could follow or recommend. The criterion of non-contradictory assertability is not arbitrary but constitutive of what it means to propose a norm at all.

The objection that ethics might be "about something else entirely" (consequences, virtues, divine commands) does not escape this analysis. Any alternative ethical framework must still be arguable. The consequentialist who says "maximize utility" must presuppose self-ownership to make the argument; the virtue ethicist who says "cultivate excellence" must presuppose the right to cultivate; the divine command theorist who says "obey God" must presuppose the right to speak and advocate. Whatever the content of one's ethical theory, the act of proposing it presupposes the Hoppean framework.

The Scope of the Argument

A potential objection asks how argumentation ethics applies to those who cannot argue: infants, the severely cognitively impaired, the temporarily unconscious. If rights derive from the presuppositions of argumentation, do non-arguers have rights?

The response is that every arguer was once such a person.[15] Anyone engaged in argumentation must value the conditions that made their current capacity possible, including not being killed during the period when they lacked argumentative capacity but possessed the potential to develop it. To argue that potential arguers

have no rights would be to contradict the conditions of one's own existence as an arguer. The preargumentation state is not outside the argument's scope but presupposed by it.

This book adopts Hoppe's argument as the normative foundation for privacy. The argument has withstood nearly four decades of critical scrutiny. Those who wish to reject it bear the burden of showing how they can do so without performative contradiction.

4.4 Implications: Property and Non-Aggression

If Hoppe's argument succeeds, a compact ethical structure follows.

Self-ownership is the first implication. Individuals have exclusive rights over their own bodies and minds. Others may not use a person's body or interfere with their mental processes without consent. Bodily integrity (freedom from assault, battery, and confinement), mental integrity (freedom from manipulation, coercion, and psychological invasion), and expressive control (the freedom to communicate or to remain silent) all follow from the same underlying claim: the exclusive right to direct one's own action.

Property rights in external resources follow next. Action requires means, and means must be allocated somehow; the only allocation rule that is not itself an aggression is original appropriation of previously unowned resources and voluntary transfer thereafter. Property rights on this account include exclusive use (the owner decides how the resource is used), transfer (the owner can give or sell the resource), and exclusion (the owner can prevent others from using the resource). These are not three separate rights but three aspects of one right, tracking the three ways another person might attempt to interfere.

The Non-Aggression Principle is the ethical framework that ties the two together. The initiation of force against persons or property is illegitimate; force is justified only in defense against prior aggression. Taking property without consent is aggression. Harming persons without consent is aggression. Using deception to obtain property or consent is aggression. Forcing someone to

reveal information about themselves violates self-ownership and is therefore aggression as well.

For privacy, the consequences are specific. Mental privacy is protected because thoughts and preferences fall under self-ownership. Communication privacy is protected because choosing what to reveal and to whom is an exercise of self-ownership over one's own expressive output. Data privacy is protected because information about oneself, stored on one's own media or shared under terms one set, falls under property rights over the media and self-ownership over the information's origin. And coerced surveillance, whether forced disclosure or monitoring without consent, violates self-ownership at the root, regardless of what is learned.

4.5 The Argumentation Axiom and Privacy

Section 4.4 set out the implications of Hoppe's argument for property and non-aggression in general terms. One implication deserves its own treatment: what the argument entails specifically for privacy.

Cognitive privacy as a presupposition of argument

Hoppe's argument begins with the observation that arguers must have exclusive control over their bodies and minds to formulate and express their positions. The mind half of that claim reaches further than it first appears.

To argue is to deliberate before speaking. A position is formed through an interior process: evaluating evidence, weighing competing claims, choosing words, deciding what to assert and what to withhold. That process is not itself the argument; it is the condition under which the argument becomes possible. An arguer whose deliberative process were open to inspection and manipulation by another party before expression would not be deliberating freely. The conclusions reached under such conditions would not be genuinely that arguer's own. They would be artifacts of the observer's access.

Coerced disclosure of unexpressed thought, and surveillance of the mental process that precedes speech, therefore violates the presuppositions of argumentation in the same way that physical coercion of the body does. The one removes control over what the arguer says; the other removes control over what the arguer thinks before saying it. Both defeat the condition the argument requires.

Cognitive privacy names this condition: the exclusive right to the deliberative process that precedes expression. It is not derived from a general right to secrecy but from what argumentation itself presupposes.

Expressive control as selective disclosure

The argumentation axiom also entails expressive control: the right to choose what to reveal and to whom, and to withhold what one has not chosen to disclose.

An arguer who lacks the right to remain silent, or who can be compelled to disclose more than the argument requires, does not have real control over expressive output. Forced disclosure is not a condition under which argument can proceed freely. The arguer who can be required to produce all internal deliberation, prior communications, correspondence, and associations is not free to argue; they are subject to a form of coercion that precedes the physical.

Eric Hughes's formulation in the Cypherpunk Manifesto[16] maps precisely onto this structure. Privacy is selective disclosure: each party to a transaction reveals what the transaction requires and nothing beyond. Hughes arrived at this through engineering instinct; Hoppe's argumentation ethics gives it a normative foundation. The right to selective disclosure is not a preference but a presupposition of rational discourse. Any system that removes it removes a condition of argumentation itself.

Surveillance as performative contradiction

A state that compels total transparency, requiring citizens to produce communications, disclose associations, reveal financial

transactions, and submit to continuous observation, is engaged in a performative contradiction of a specific kind.

The state must justify its authority through argument. It passes laws, issues rulings, makes claims on citizens, and invites challenge in courts. Every one of these activities is a form of argumentation, and every one presupposes the framework Hoppe identifies: the participants must have exclusive control over their own deliberative processes and expressive outputs to engage in actual discourse about the norms that govern them.

A state that removes cognitive privacy and expressive control from its subjects has destroyed the conditions under which those subjects could participate as real parties to any normative discourse. It has produced a population that cannot argue back, because the prerequisites of argument have been stripped away. The surveillance regime is not just an aggression against individual rights; it is an assault on the institutional conditions of discourse itself.

From norm to architecture

What the argumentation axiom establishes is a norm, not a mechanism. It tells us that cognitive privacy and expressive control are presuppositions of rational discourse, and that their violation is aggression in the full Hoppean sense. What it does not tell us is how those presuppositions can be protected against a surveillance apparatus that does not care whether its subjects can argue back.

That question is where the cypherpunk tradition takes over from the normative one. Chapter 5 introduces the Axiom of Resistance, which translates the norm into a cost structure. The chapters that follow show which cryptographic and economic tools raise the cost of observation enough to make the norm enforceable in practice.

Chapter Summary

Hoppe's argumentation ethics shows that engaging in discourse presupposes self-ownership: the denier of self-ownership must exercise exclusive control over body and mind to formulate and express the denial, which is performative contradiction. Property rights

follow from self-ownership through original appropriation, the only universalizable rule for resolving conflicts over scarce resources. The Non-Aggression Principle follows in turn: uninvited interference with person or property is aggression, and force is justified only in defense against it. Coerced surveillance violates self-ownership and therefore counts as aggression by the same framework.

The argument has faced objections over four decades and has answered them. Murphy and Callahan asked whether Hoppe establishes ownership or mere use; the distinction collapses under scrutiny, and refusing to acknowledge ownership places one outside rational discourse entirely. Hume's is-ought question asks whether the argument bridges description and prescription; it does not derive ought from is but shows that certain normative claims cannot be coherently denied. A partial-application objection asks why principles established in argumentation apply outside it; universalizability is constitutive of argumentation itself. Kinsella's estoppel argument, the preargumentation defense for potential arguers, and van Dun's clarification that rejecting the argument's presuppositions places one outside the community of discourse all strengthen the same framework.

The argumentation axiom reaches further into privacy than the general aggression analysis alone. Argumentation presupposes cognitive privacy: the deliberative process that precedes expression must be exclusively controlled by the arguer, or the conclusions reached are not that arguer's own. It also presupposes expressive control: the right to choose what to disclose and to withhold what one has not offered. These are not derived from a general right to secrecy but from what discourse requires. A state that strips cognitive privacy and compels total disclosure has destroyed the conditions under which its subjects could participate as real parties to any normative argument about the rules that govern them. The surveillance regime is a performative contradiction at the institutional level: it claims to rule by law and argument while removing the presuppositions that make law and argument possible.

What this provides is the criterion, not a complete rulebook. Consent, public observation, and inference from available informa-

tion each require case-by-case application of the principle. The framework also does not establish that privacy is enforceable in practice. Chapter 5 takes up the Axiom of Resistance; the technical and economic chapters that follow turn the norm into something the world can maintain.[17]

Notes

[1]Hans-Hermann Hoppe, "Argumentation Ethics," in *The Economics and Ethics of Private Property*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2006), 380.

[2]Hans-Hermann Hoppe, *A Theory of Socialism and Capitalism: Economics, Politics, and Ethics* (Boston: Kluwer Academic Publishers, 1989; repr. Auburn, AL: Ludwig von Mises Institute, 2010), `https://mises.org/library/theory-socialism-and-capitalism`. Chapter 7 contains the full original statement of argumentation ethics from which this chapter's analysis derives; subsequent chapters apply the argument to derive private property norms and a systematic critique of socialist and interventionist arrangements.

[3]On discourse ethics, see Jürgen Habermas, *Moral Consciousness and Communicative Action*, trans. Christian Lenhardt and Shierry Weber Nicholsen (Cambridge, MA: MIT Press, 1990); Karl-Otto Apel, "The Problem of Philosophical Foundations in Light of a Transcendental Pragmatics of Language," in *After Philosophy: End or Transformation?*, ed. Kenneth Baynes et al. (Cambridge, MA: MIT Press, 1987). For Hoppe's adaptation of discourse ethics into argumentation ethics, see Chapter 3, note 7.

[4]John Locke, *Two Treatises of Government* (1689), Second Treatise, ch. V, §27: "Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his labour with, and joined to it something that is his own, and thereby makes it his property." Locke's labor-mixing theory is the historical origin of original appropriation as a property-generating act; Hoppe's homesteading principle inherits this structure while grounding it in the presuppositions of argumentation rather than in natural theology.

[5]Murray N. Rothbard, *The Ethics of Liberty* (New York: NYU Press, 1998; original Humanities Press, 1982), `https://mises.org/library/ethics-liberty`. The canonical libertarian statement of the Non-Aggression Principle, grounded in self-ownership: "No man or group of men may aggress against the person or property of anyone else." Hoppe's argumentation ethics is designed to provide this Rothbardian NAP with a firmer, non-circular foundation than Rothbard's own natural-rights derivation.

[6]Robert P. Murphy and Gene Callahan, "Hans-Hermann Hoppe's Argumentation Ethic: A Critique," *Journal of Libertarian Studies* 20, no. 2 (2006): 53–64. Available at `https://cdn.mises.org/20_2_3.pdf`.

[7]For responses to Murphy and Callahan, see Walter Block, "Rejoinder to Murphy and Callahan on Hoppe's Argumentation Ethics," *Journal of Libertarian Studies* 22 (2010): 631–639; Marian Eabrasu, "A Reply to the Current Critiques Formulated Against Hoppe's Argumentation Ethics," *Libertarian Papers* 1, no. 20 (2009).

[8]Frank van Dun, "Argumentation Ethics and the Philosophy of Freedom," *Libertarian Papers* 1, no. 20 (2009). Van Dun's analysis clarifies that one who refuses to acknowledge the presuppositions of argumentation has not refuted the argument but placed themselves outside the community of rational discourse.

[9]Hatim Kheir, "Bridging the Use-Ownership Gap: A Reformulation of Hoppe's Argumentation Ethics on Praxeological Grounds," *Journal of Libertarian Studies* 27, no. 1 (2023): 115–134. Available at `https://jls.mises.org/article/84874`.

[10]On universalizability, see Hoppe, "Argumentation Ethics," 383–384; Stephan Kinsella, "Explaining Argumentation Ethics and Universalizability Concisely," (2019), available at `https://stephankinsella.com/2019/03/explaining-argumentation-ethics-and-universalizability-concisely-to-a-facebook-friend/`. The argument that particularistic norms are self-defeating in rational discourse derives from the Kantian tradition but receives distinctly libertarian application in Hoppe's framework.

[11]On estoppel as a complementary justification for rights, see Stephan Kinsella, "Estoppel: A New Justification for Individual Rights," *Reason Papers* 17 (1992): 61–74. Kinsella argues that aggressors are estopped from objecting to defensive force because their own actions demonstrate acceptance of the principle that force is permissible.

[12]David Hume, *A Treatise of Human Nature* (1739–40), book III, part I, §1. The is-ought distinction appears in Hume's observation that authors move imperceptibly from "is" and "is not" to "ought" and "ought not"; he flagged the inferential gap without resolving it.

[13]See Kinsella's interpretation of Hoppe's structure: Stephan Kinsella, "Argumentation Ethics and Liberty: A Concise Guide," Mises Institute (2011). Available at `https://mises.org/mises-daily/argumentation-ethics-and-liberty-concise-guide`.

[14]Murray N. Rothbard, "Beyond Is and Ought," *Liberty* 2, no. 2 (1988): 44–45. Rothbard stated that Hoppe "has managed to transcend the famous is/ought, fact/value dichotomy that has plagued philosophy since the days of the scholastics."

[15]On the extension of argumentation ethics to potential arguers, see Patrick Tinsley, "Preargumentation Ethics and the Issue of Abortion," *Journal of Libertarian Studies* (2020). The argument holds that current arguers must value the conditions that made their own argumentative capacity possible, including protection during the period when they possessed potential but not actual argumentative capacity.

[16]Eric Hughes, "A Cypherpunk's Manifesto" (1993), available at `https://www.activism.net/cypherpunk/manifesto.html`. Hughes

defines privacy as selective disclosure: "Privacy is the power to selectively reveal oneself to the world." His manifesto is the founding document of the cypherpunk movement and the direct source of the selective-disclosure formulation cited in §4.5. Hoppe's argumentation ethics gives Hughes's engineering intuition a normative foundation: the right to selective disclosure is a presupposition of rational discourse, not a preference.

[17]Further reading on libertarian ethics and argumentation ethics. For the primary Hoppe texts (*A Theory of Socialism and Capitalism*, ch. 7, and *The Economics and Ethics of Private Property*), see Chapter 3, note 7. For the Rothbardian natural-rights tradition that Hoppe extends, Murray N. Rothbard, *The Ethics of Liberty* (NYU Press, 1998; original 1982), is the canonical source. For the shorter entry, Stephan Kinsella's ongoing writing at `https://stephankinsella.com` systematically unpacks Hoppe's arguments and responds to critics; his *Legal Foundations of a Free Society* (Papinian Press, 2023) collects his treatments. On the broader tradition of natural-rights libertarianism, see Robert Nozick, *Anarchy, State, and Utopia* (Basic Books, 1974), and Eric Mack, *Libertarianism* (Polity, 2018). For the analytic-philosophy treatment of the is-ought problem that argumentation ethics addresses, see J. L. Mackie, *Ethics: Inventing Right and Wrong* (Penguin, 1977), and the *Stanford Encyclopedia of Philosophy* entry on "Moral Anti-Realism" at `https://plato.stanford.edu/entries/moral-anti-realism/`. For the Habermasian "discourse ethics" that Hoppe's argumentation ethics partially parallels (and critiques), see Jürgen Habermas, *Moral Consciousness and Communicative Action*, trans. Christian Lenhardt and Shierry Weber Nicholsen (MIT Press, 1990), and Karl-Otto Apel, *Understanding and Explanation: A Transcendental-Pragmatic Perspective*, trans. Georgia Warnke (MIT Press, 1984). For critical appraisals of Hoppe's argument specifically, Robert P. Murphy and Gene Callahan, "Hans-Hermann Hoppe's Argumentation Ethic: A Critique," *Journal of Libertarian Studies* 20, no. 2 (2006): 53–64; and the replies in subsequent issues of the same journal, catalog the main objections. For a self-ownership case independent of Hoppe, Michael Huemer, *The Problem of Political Authority* (Palgrave Macmillan, 2013), grounds strong property rights in common-sense moral intuitions.

5

THE AXIOM OF RESISTANCE

> "One who does not accept the axiom of resistance is contemplating an entirely different system than Bitcoin."
>
> Eric Voskuil, *Cryptoeconomics* (2019)[1]

The book uses "axiom" three times: the Action Axiom of Chapter 3, the Argumentation Axiom of Chapter 4, and the Axiom of Resistance of this chapter. The three are not axioms in the same sense, and the difference is worth marking before the chapter proceeds.

The Action Axiom is self-evident in the strict praxeological sense. Mises holds that the proposition "human beings act purposefully" cannot be denied without performing the purposeful behavior the denial would deny. It is true a priori and grounds the descriptive science of action.

The Argumentation Axiom is performatively undeniable in Hoppe's sense. Anyone who enters argument has already presupposed self-ownership, the right to use one's body to make claims, and the norms that make discourse possible. It grounds the normative side of the framework.

The Axiom of Resistance is methodological. It is a working assumption about engineering possibility, drawn from cryptographic theory and the empirical record of deployed systems, that defines the subject matter of any analysis of freedom technology. Eric Voskuil introduced the term in *Cryptoeconomics* to mark the boundary between systems that can resist external control (Bitcoin, Tor, end-to-end encryption) and systems that cannot (PayPal, central-

bank money, platform-mediated identity). One who rejects this assumption commits no logical error; the rejection shifts the subject matter to systems that do not resist.

The three serve different roles. Action explains the economics of voluntary order. Argumentation supplies the ethics that prohibits aggression. Resistance supplies the strategy without which the first two cannot be defended in practice. Perfect ownership of goods is an empty norm if the owner hands the goods over the moment an aggressor appears. The three together form the backbone of any thorough analysis of freedom technology, and the rest of the chapter examines what the third assumption asserts and on what grounds it is well-supported.

This chapter presents the third foundation: the Axiom of Resistance. Unlike the preceding axioms, this is an assumption, not a self-evident or normative claim. Eric Voskuil, in his analysis of Bitcoin's security model, states that it is "not accepted as a fact but deemed a reasonable assumption, due to the behavior of similar systems."[1]

The assumption is this: systems can be designed to resist external control. Cryptographic tools can make surveillance technically difficult. Mathematical properties can protect privacy in ways that political promises cannot. You can coherently deny this assumption; PayPal-type systems, which rely on central authority, are perfectly coherent objects of analysis. Rejecting the Axiom of Resistance does not commit any logical error. It shifts the subject matter to the design of systems that do not resist.

But the assumption is well-grounded. It rests on mathematical foundations and empirical evidence, reinforced by methodological necessity. This chapter examines what the axiom asserts and why it is well-grounded, then marks where its limits begin.

5.1 What the Axiom Asserts

The Core Claim

The Axiom of Resistance asserts that it is possible to design systems that resist external control, and it rests on several concrete

capabilities. Cryptographic systems can protect information from unauthorized access, decentralized networks can operate without single points of control, and mathematical properties can provide stronger guarantees than legal or political protections. Resistance is not guaranteed, but it is technically achievable.

What ``Resistance'' Means

Resistance is the capacity to impose costs on adversaries attempting control. A system resists to the degree that circumventing its protections requires resources exceeding what adversaries are willing or able to expend. Resistance is not binary but exists on a spectrum: a system may resist casual attackers but not nation-states, or resist all known attacks but remain vulnerable to advances in mathematics or computing.

Resistance operates across several dimensions: computational (cryptographic systems require resources that are practically unobtainable, as a 256-bit key demands more brute-force operations than the age of the universe permits), economic (the cost of attack exceeds the value of success, so even a breakable system resists if breaking it costs more than the information is worth), structural (decentralized systems present no single point to attack, since taking down one node leaves others operating and compromising one participant does not compromise the network), and jurisdictional (systems distributed across legal boundaries resist because no single authority can compel compliance from all components).

Resistance is not invulnerability. No system is perfectly secure. The axiom claims that resistance is possible: systems can be designed such that overcoming them requires resources exceeding what attackers are willing to expend. Resistance is also asymmetric, meaning defenders can achieve protection at lower cost than attackers can achieve breach; encryption is cheap, while breaking strong encryption is expensive. This asymmetry shifts the balance of power. Without cryptographic protection, surveillance is easy and privacy is hard. With cryptographic protection, privacy becomes feasible and surveillance becomes costly.

What the Axiom Does NOT Assert

The axiom does not claim that resistance is guaranteed; systems can fail, implementations can have bugs, users can make mistakes, so the axiom assumes resistance is possible, not that it always succeeds. Nor does it claim that resistance is absolute, since states have resources individuals lack and can sometimes overcome resistance through legal compulsion, physical coercion, or massive resource expenditure when sufficiently motivated. The axiom does not claim that resistance solves all problems, as physical coercion, social engineering, and human error affect both costs and outcomes. Finally, unlike the Action Axiom, denial of resistance creates no logical contradiction; the axiom is assumed, not proven, and this distinguishes its logical status from the self-evident foundations examined in earlier chapters.

5.2 Why the Axiom Is Well-Grounded

Though an assumption and not a proof, the Axiom of Resistance rests on substantial foundations.

Mathematical Grounding

Modern cryptography rests on computational hardness assumptions.[2] Certain mathematical problems appear to be difficult to solve. Factoring large numbers presents one such challenge: given two large primes, multiplying them is easy, but given their product, finding the original primes is computationally infeasible with current technology. The discrete logarithm problem exhibits similar properties; in certain mathematical structures, computing a value is easy while reversing the computation is infeasible.[3] Hash function preimage resistance provides another foundation: given a cryptographic hash output, finding an input that produces that output is computationally infeasible. These hardness assumptions underlie RSA,[4] elliptic curve cryptography,[5] and the hash functions used in Bitcoin and other systems. If the assumptions hold, the cryptographic protections are real.

The assumptions are not proven. If P equals NP, a question that remains open, most current cryptographic assumptions would collapse. The entire edifice of public-key cryptography rests on conjectures that, while well-supported by decades of failed attacks, have no mathematical proof of correctness.

Algorithmic progress continues. The General Number Field Sieve[6] improved factoring efficiency over earlier methods, lattice-based attacks weakened certain elliptic curve implementations, and quantum computing, discussed in Chapter 14, threatens to break most current public-key cryptography entirely.[7] The security margins that seem comfortable today may narrow as mathematics and computing advance.

What decades of research have established is not that these problems are provably hard, but that no one has yet found efficient solutions. The assumptions are empirically well-grounded, not mathematically proven. Cryptographic security is therefore contingent on the continued failure of attack research, not on impossibility already proved. Chapter 14 examines these computational foundations in detail.[8]

Empirical Track Record

Systems designed for resistance have shown actual resistance. The Tor network, operating since 2002, has provided anonymous communication at meaningful scale despite state-level adversaries. Its limits are real: timing attacks, compromised exit nodes,[9] and user error remain concerns. Even so, Tor has continued operating and remains one of the most durable public anonymity systems ever deployed. Bitcoin, operating since 2009, has processed transactions and maintained consensus without a central authority; states have constrained exchanges, miners, and access points, but they have not shut down the network. PGP and its descendants have protected communications for decades with strong encryption.[10] When keys are unavailable, prosecution still does not give authorities decryption by fiat. Signal and other end-to-end encrypted messaging applications have protected private communications at scale.[11]

This empirical record does not prove resistance will always succeed. But it establishes that resistance has succeeded in practice over extended periods against well-resourced adversaries.

Methodological Necessity

Voskuil emphasizes that accepting the Axiom of Resistance defines what we are analyzing. "One who does not accept the axiom of resistance is contemplating an entirely different system than Bitcoin."

If you assume resistance is impossible, you are analyzing permissioned systems: PayPal, bank accounts, regulated financial institutions. These systems operate at the pleasure of authorities and can be shut down or surveilled at will.

If you assume resistance is possible, you are analyzing permissionless systems: Bitcoin, Tor, end-to-end encryption. These systems operate independently of authorities and resist control by design.

Both types of systems exist. Both are worthy of study. But they are different, and analysis appropriate to one may not apply to the other. The axiom is a methodological choice that defines the subject matter.

Epistemic Problem with Denial

There is an epistemic peculiarity in denying the Axiom of Resistance.

If resistance is impossible, if states can control all information flows and overcome all cryptographic protection, then how would you know? Your sources of information would be controlled. Your ability to discover resistance possibilities would be limited. Your belief that resistance is impossible might be a product of the control you think is total.

The observation is not a proof that resistance is possible. But it suggests that confident denial faces its own epistemic challenges. The denier cannot easily verify their denial without access to information that, if the denial is correct, they cannot trust.

5.3 Relationship to Other Foundations

Action Axiom (Chapter 3)

The Action Axiom establishes that privacy is built into the structure of action. Deliberation is internal; preferences are subjective; information asymmetry is inherent.

The Axiom of Resistance asks: can this inherent property be protected? Can the privacy that exists as a fact of human action be preserved against attempts to eliminate it?

The Resistance Axiom assumes yes. Technical protection is possible. The structural privacy of action can be maintained through cryptographic means.

Argumentation Axiom (Chapter 4)

The Argumentation Axiom argues that privacy cannot be coherently denied in discourse. Self-ownership includes control over one's mental processes and communications.

But this normative conclusion is empty without implementation. Claiming that privacy should be protected does not make it protected.

The Axiom of Resistance bridges the gap between normative and practical. If resistance is possible, then the privacy that ought to be protected (per Chapter 4) can be protected (per Chapter 5).

The Three Foundations Together

The three foundations fit together. Chapters 3 and 4 established what privacy is and grounded its moral and economic stakes. This chapter adds the practical premise: it can be protected.

Chapter Summary

The Axiom of Resistance is the third foundation: systems can be designed to resist external control, not as a proved fact but as a well-grounded assumption. It rests on computational hardness in cryptography, the empirical record of Tor,[12] Bitcoin,[1] and end-to-

end encryption resisting pressure for years, and the methodological point that accepting the axiom defines the subject matter of the rest of the book. The epistemic peculiarity of its denial, that a person who could never verify resistance also cannot reliably conclude it is impossible, does not prove the axiom but narrows what confident rejection can coherently assert.

The axiom claims possibility, not inevitability. Resistance often fails under physical coercion (the "$5 wrench attack"),[13] implementation error,[14] user mistakes, or insufficient network scale. It is also not costless. Cryptography makes digital defense cheap and digital attack expensive, but physical coercion inverts the asymmetry, and system design can shift costs without eliminating them. Today's security may be tomorrow's vulnerability as threats evolve and states bring resources individuals lack: legal authority, compelled cooperation, supply-chain access. The assumption remains an assumption, and reasonable people can reject it while analyzing different systems. What it makes possible is the rest of the book, which proceeds under the first choice.[15]

Notes

[1]Eric Voskuil, "Axiom of Resistance," *Cryptoeconomics: Fundamental Principles of Bitcoin*, available at `https://github.com/libbitcoin/libbitcoin-system/wiki/Axiom-of-Resistance`.

[2]On computational hardness and cryptographic assumptions, see Jonathan Katz and Yehuda Lindell, *Introduction to Modern Cryptography*, 3rd ed. (Boca Raton: CRC Press, 2020), particularly chapters on computational security and hardness assumptions.

[3]On the discrete logarithm problem as a cryptographic hardness assumption, see Whitfield Diffie and Martin E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* 22, no. 6 (1976): 644–654. This foundational paper introduced public-key cryptography by exploiting the asymmetry between computing discrete logarithms and inverting them, establishing the computational basis that RSA and elliptic curve systems later extended.

[4]R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM* 21, no. 2 (1978): 120–126. The original RSA paper; the system's security rests on the integer factorization hardness assumption described in the surrounding paragraph.

[5]Victor S. Miller, "Use of Elliptic Curves in Cryptography," in *Advances in Cryptology — CRYPTO '85*, Lecture Notes in Computer Science 218 (Berlin: Springer, 1986), 417–426; Neal Koblitz, "Elliptic Curve Cryptosystems," *Mathematics of Computation* 48, no. 177 (1987): 203–209. These two independent papers introduced elliptic curve cryptography; ECC achieves equivalent security to RSA with shorter keys, making it the dominant public-key primitive in modern deployments including Bitcoin's secp256k1 curve.

[6]A. K. Lenstra, H. W. Lenstra Jr., M. S. Manasse, and J. M. Pollard, "The Number Field Sieve," in *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing* (1990), 564–572. The General Number Field Sieve is the most efficient known classical algorithm for factoring large integers, running in sub-exponential time; its development required significant upward revision of RSA key-length recommendations and illustrates how algorithmic progress can erode assumed security margins.

[7]On post-quantum cryptography, see NIST Post-Quantum Cryptography Standardization project, `https://csrc.nist.gov/projects/post-quantum-cryptography`. NIST released the first finalized post-quantum encryption standards in August 2024.

[8]On the P vs NP problem and its implications for cryptography, see Scott Aaronson, *Quantum Computing Since Democritus* (Cambridge: Cambridge University Press, 2013), and Michael Sipser, *Introduction to the Theory of Computation*, 3rd ed. (Boston: Cengage Learning, 2012).

[9]On Tor's known attack surface, see Steven J. Murdoch and George Danezis, "Low-Cost Traffic Analysis of Tor," *Proceedings of the 2005 IEEE Symposium on Security and Privacy* (2005), 183–195; and Aaron Johnson et al., "Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries," *Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications Security* (2013), 337–348. These papers document timing-correlation and exit-node attacks that remain the primary practical limitations of the Tor anonymity model against well-resourced adversaries.

[10]Phil Zimmermann released PGP (Pretty Good Privacy) in 1991; for "Why I Wrote PGP" (1991) and the cypherpunk-history treatment, see Chapter 2, note 15. The protocol is standardized as OpenPGP, currently RFC 9580 (July 2024) with RFC 4880 (2007) as the historical reference; the canonical project page is `https://www.openpgp.org/`.

[11]Signal Foundation, `https://signal.org/`, with protocol specifications at `https://signal.org/docs/`. The Double Ratchet protocol underlying Signal is the reference design for modern end-to-end encrypted messaging; for the wider systems-cluster treatment, see Chapter 2, note 8. Substantive treatment of Signal occurs in Chapter 17.

[12]On Tor's security model and limitations, see Roger Dingledine, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router," USENIX Security Symposium (2004). For ongoing analysis, see `https://blog.torproject.org/`.

[13]Randall Munroe, "Security," *xkcd* #538, https://xkcd.com/538/. The comic depicts an attacker bypassing strong encryption by physical coercion ("hit him with this $5 wrench until he tells us the password"), the canonical illustration that cryptographic resistance does not extend to the threat model of physical force. The phrase "rubber-hose cryptanalysis" is the older synonym in the cryptographic literature.

[14]On implementation failures and their lessons, see Ross Anderson, *Security Engineering*, 3rd ed. (Indianapolis: Wiley, 2020), particularly the chapters on cryptographic implementation failures.

[15]Further reading on resistance, security, and the cryptographic-political interface. On the theoretical foundations of computational security, Katz and Lindell, *Introduction to Modern Cryptography* (cited in note 3 above), is the graduate-level standard. For the practitioner introduction (Ferguson, Schneier, and Kohno) and the canonical treatment of cryptography and state capability (Diffie and Landau, *Privacy on the Line*; Landau, *Listening In*), see Chapter 1, note 5. For the conceptual case that cryptography changes political economy, Timothy C. May's *Cyphernomicon*, already cited, is the primary-source sprawling cypherpunk framing. For the cryptocurrency-specific version, see Voskuil, *Cryptoeconomics* (cited in note 1 above), and the wider Austrian-praxeological treatment at Chapter 2, note 5. On resistance as a broader phenomenon beyond cryptography, James C. Scott's corpus is the standard: for *Seeing Like a State* (Yale University Press, 1998), see Chapter 1, note 6; the wider corpus extends through *Weapons of the Weak: Everyday Forms of Peasant Resistance* (Yale University Press, 1985) and *The Art of Not Being Governed: An Anarchist History of Upland Southeast Asia* (Yale University Press, 2009). For Austrian-libertarian political theory on resistance, Murray N. Rothbard, *For a New Liberty: The Libertarian Manifesto* (Macmillan, 1973), and *The Ethics of Liberty* (cited at Chapter 2, note 3); and Hans-Hermann Hoppe, *Democracy: The God That Failed: The Economics and Politics of Monarchy, Democracy, and Natural Order* (Transaction, 2001), are the canonical texts. For the analytic defense of philosophical anarchism from outside the Austrian tradition, Michael Huemer, *The Problem of Political Authority* (cited at Chapter 4, note 13). On the post-quantum transition and its strategic implications, Michele Mosca, "Cybersecurity in an Era with Quantum Computers: Will We Be Ready?" *IEEE Security & Privacy* 16, no. 5 (2018): 38–41, https://arxiv.org/abs/1512.04202, introduces the "Mosca inequality" framework; the ongoing NIST process at https://csrc.nist.gov/projects/post-quantum-cryptography tracks the standardization.

PART III

ECONOMIC FOUNDATIONS

6

INFORMATION, SCARCITY, AND PROPERTY

> "If nature has made any one thing less susceptible than all others of exclusive property, it is the action of the thinking power called an idea, which an individual may exclusively possess as long as he keeps it to himself; but the moment it is divulged, it forces itself into the possession of every one, and the receiver cannot dispossess himself of it."
>
> Thomas Jefferson, letter to Isaac McPherson (1813)[1]

What is the relationship between information and property? The answer determines how privacy can be protected. If information content can be property, privacy might be protected through property rights in information itself. If information content cannot be property, privacy must be protected through other means.

Stephan Kinsella's analysis provides the framework.[2] Property rights apply only to scarce resources, those that one party's use precludes another's use. Information content, once known, can be used by unlimited parties simultaneously without depletion. It is non-scarce and therefore cannot be property.

This chapter follows the Kinsella framework because it is the direct application of scarcity-based property theory that Austrian economics has developed from Rothbard forward. The framework is not the only Austrian-libertarian position on intellectual property,

and the chapter says so plainly. Murray Rothbard, in *Man, Economy, and State*, treated patent and copyright as separate cases: patent grants a monopoly that suppresses independent inventors and is illegitimate; copyright restated as a contractual reservation by an author against purchasers (the original meaning before the modern statutory regime) is in principle legitimate as a contract, though it does not bind third parties who never agreed to it. Kinsella's later development sharpens Rothbard's analysis by showing that the contractual route does not in fact reach what copyright statute does in operation (binding non-parties, controlling downstream physical media, surviving into the public domain), and by extending the same scarcity-based logic to all categories of state-granted information monopoly. The two positions converge on the points that matter for privacy: self-ownership covers the body and mind, physical property covers devices and premises, and contract creates enforceable confidentiality between parties who agree to it.[3] The chapter develops the Kinsella version because it is more precise about what contract can and cannot do, but the privacy argument the book defends survives in either form.[4]

This does not leave privacy unprotected. Privacy is protected through self-ownership (control over one's body and mind), physical property (devices, papers, homes), and contract (confidentiality agreements). These mechanisms protect privacy without treating information content as property, and understanding this distinction is essential for clear analysis of privacy economics.

6.1 Scarcity as the Foundation of Property

Why Property Rights Exist

Property rights exist to resolve conflicts over scarce resources. A scarce resource is one that, when used by one party, cannot simultaneously be used by another in the same way. Land, food, tools, and bodies are scarce: my use of this apple precludes your simultaneous use of the same apple.

Scarcity creates the potential for conflict. If multiple parties want to use the same resource in incompatible ways, they must

either fight over it or agree on allocation rules. Property rights are allocation rules that assign control over resources to specific parties. Economists call this property of physical goods "rivalrousness": one party's use excludes another's. Information is non-rivalrous — when you learn something I know, my knowledge is not reduced — which is why the same property logic does not apply to it.

Without scarcity, property rights serve no purpose. If a resource can be used by everyone simultaneously without conflict, no allocation rules are needed.[5]

A related distinction makes the framework cleaner. Possession is a factual matter: who currently controls a resource. Ownership is a normative one: who has the legal right to control it. Roman law and the Anglo-American common law both recognize the distinction, and Kinsella's treatment makes it foundational.[6] Property rights provide normative support for possession; they answer the question of who, in a dispute, has the better claim. Possession is descriptive. Ownership is prescriptive. The acquisition rules below allocate ownership; they do not describe who happens to be holding the thing at any given moment.

Physical Scarcity vs. Artificial Scarcity

Physical scarcity is inherent in the resource itself. Land is scarce because occupying one location precludes simultaneous occupation by another. Apples are scarce because eating one prevents anyone else from eating the same apple.

Artificial scarcity is imposed by external force on resources that are not inherently scarce. If I write a poem and you memorize it, you can recite it without diminishing my ability to recite it. The poem, as a pattern of words, is non-scarce. But if the state grants me a "copyright" enforced by violence, it artificially restricts your use of a non-scarce resource.

Property rights in physically scarce resources resolve real conflicts. "Property rights" in artificially scarce resources create conflicts that would not otherwise exist. They give one party control over how others may use their own legitimately owned physical resources.

6.2 ``Intellectual Property'' as Aggression

Section 6.1 established that property rights exist to resolve conflicts over scarce resources, and that information content is non-scarce. The implication: "intellectual property" creates artificial scarcity through state violence, violating actual property rights in physical resources.

Consider a simple case. Alice invents a new mousetrap design. Bob independently develops the same design, or learns of it and implements it using his own materials.

If Alice has "intellectual property" in the design, she can use state violence to prevent Bob from using his own materials (wood, springs, wire) in certain configurations. The state will punish Bob for arranging his own property in ways Alice disapproves.

The claim constitutes aggression. Bob has done nothing to Alice or her property. He has used his own property. The "intellectual property" claim is a claim that Alice can control how Bob uses Bob's property, backed by state violence.

The Patent Example

Patents grant monopolies on ideas. A patent holder can prevent anyone else from implementing the patented idea, even someone who invented it independently.

The patent holder controls how others may use their own physical property. Independent inventors have no defense because the first to file wins. Competition is restricted not by superior service but by legal privilege. Innovation is taxed, as anyone improving on patented ideas must pay tribute.

Patents are state-granted monopoly privileges enforced through aggression against others' actual property.[7]

The Copyright Example

Copyright grants control over copying patterns. A copyright holder can prevent others from arranging their own physical property (paper, ink, hard drives) in certain configurations.

You cannot print certain patterns on your own paper, store certain bit patterns on your own hard drive, or speak certain word sequences in commercial contexts. Your physical property is controlled by others' claims to patterns.

Copyright operates as censorship backed by state violence: it restricts what you may do with your own physical property based on pattern similarity to patterns someone else claims.

IP as Nonconsensual Negative Servitude

The legal name for what the state grants the IP holder makes the framework precise. A negative servitude (or easement) is a property interest that lets one party prevent the owner of another resource from using it in specified ways. Consensual negative servitudes are common and unobjectionable: a restrictive covenant in a residential subdivision that forbids commercial use is a negative servitude every buyer accepts when buying in. What patent and copyright create is a *nonconsensual* negative servitude, granted by the state without the agreement of the owner whose use is restricted. The independent inventor never agreed to refrain from arranging his own materials in the patented configuration. The reader who buys a book never agreed to refrain from arranging bits on his own hard drive in the copyrighted pattern. The IP holder acquires, by state grant, the power of a co-owner over property he never owned and never bought.[8]

Creation Is a Source of Wealth, Not of Property

The error that keeps recurring under all this is the Lockean intuition that creation is a source of property rights, the idea that mixing one's labor with something gives one ownership of it. Creation is a source of *wealth*, not a source of *property*. Every act of creation transforms already-owned input materials into a more valuable output configuration; the owner of the inputs is the owner of the outputs because he was the owner of the inputs, not because he created the outputs. Employees in a factory create the products and do not own them. Original appropriation, contractual transfer,

and rectification are the three sources of property rights in scarce external resources. Self-ownership of the body grounds rights in one's person. Creation is not a fourth source. The intuition that creators ought to own their creations is the ghost that animates IP advocacy, and it has to be rejected explicitly: useful ideas, original expressions, and recognizable brands are products of creation that increase wealth in the world, and none of them is property.[9]

6.3 Content vs. Media: The Critical Distinction

Information Exists on Physical Media

Information is always instantiated on physical media: a book is paper and ink arranged in patterns, a hard drive contains magnetic domains in specific configurations, and a brain contains neural patterns encoding memories and knowledge.

The media (paper, hard drive, brain) are scarce physical objects. The content (patterns, information, ideas) is non-scarce.

Property Rights Apply to Media, Not Content

You own your paper. You can do anything with it: write on it, burn it, fold it into a hat. Your property rights are complete.

You own your hard drive. You can store any bit patterns you want. Your property rights include determining what configurations your property takes.

But once you communicate content to another person, you cannot control what they do with their own media. If you tell me a secret and I write it in my notebook, you have no property claim to my notebook. The content, now in my mind and my notebook, is not your property.

What ``Owning Information'' Would Mean

If information content were property, teaching would be transfer of property (do teachers lose their knowledge when students learn?), learning would be acquisition of property (from whom? with what consent?), memory would be storage of others' property

(can they demand deletion?), and conversation would be property exchange (tracking who "owns" each idea discussed?).

These absurdities reveal that information-as-property is incoherent. Knowledge is not a thing that can be owned, lost, stolen, or transferred in the way physical objects can.

The ``Theft'' Confusion

When someone copies your file without permission, have they "stolen" it? No. You still have your file, nothing has been taken from you, and your property is intact.

What has happened is that they have created new patterns on their own media, patterns similar to patterns on your media. This may violate contract if they had agreements with you. It may be wrong for other reasons. But it is not theft because nothing was taken.

The language of "stealing" ideas confuses the issue by importing property concepts where they do not apply.

6.4 How Privacy Is Protected

If information content cannot be property, how is privacy protected? Through three mechanisms that do not require information-as-property claims.

Self-Ownership

Chapter 4 argued that self-ownership follows from argumentation ethics. Self-ownership grounds claims to mental privacy and bodily integrity, as well as control over what you express. Your thoughts are yours, and extracting them without consent stands in tension with that claim. Your body is yours, and uninvited examination or monitoring stands in the same relation. You choose what to communicate, and silence remains within your control.

Self-ownership protects privacy at the source. Before information is communicated, it exists only in your mind and body, which are yours.

Physical Property

You own your devices and papers, your home and hard drives. Property rights over these objects include deciding who may enter and refusing search without consent. You determine how your property is configured, including what data your devices store.

Physical property rights protect information by protecting the media on which it exists. Your encrypted hard drive is your property. Others have no automatic right to access it, and any claim to compel decryption would need separate justification beyond the property claim itself.

Contract

Voluntary agreements can create enforceable obligations.[10] Non-disclosure agreements bind parties not to reveal certain information. Employment or service contracts may include confidentiality clauses. Attorneys, doctors, and clergy operate under traditional professional privilege requiring confidentiality.

Contract protects privacy through voluntary commitment. When someone agrees to keep information confidential, they become bound by that agreement.

Breach of confidentiality agreements is more than "breaking a promise." When payment is made under the condition of secrecy, revealing the secret can amount to taking payment under false pretenses. The money was transferred conditionally; accepting payment while violating the condition is at least fraud, and in some cases can be treated as theft of a conditional transfer, not mere breach of an abstract obligation. Contract violations have teeth because they involve transfers of scarce resources made under specific conditions.

Contract binds the parties who agreed to it. It does not bind third parties who never agreed. An employee under a non-disclosure agreement is bound by it; a stranger who finds the document on a sidewalk has no obligation under it. A licensee who accepts terms of use is bound on those terms; a friend who borrows the licensee's copy is bound only by whatever separate agreement, if any, the friend made. The reach of contract is the reach of consent,

and consent does not propagate to non-signatories. This is what distinguishes the contract route from a property claim in the information itself: a property claim would reach everyone, including independent discoverers and downstream non-parties; a contract reaches only those who agreed. The book treats this limit as a feature of the framework, an honest accounting of what voluntary agreement can and cannot do.

Trade Secrets

The trade secret is the cleanest case for showing what the framework does and does not protect, because it isolates the question from the patent and copyright debates.

Consider a beverage firm with a recipe its competitors would value. The recipe is protected by a combination of measures: the document itself sits in physical security; employees with access have signed non-disclosure agreements; the firm has invested heavily in vault construction and personnel screening; and the firm trades on the brand and the manufacturing capacity it has built around the secret. None of these measures involves a property claim in the recipe pattern itself. They are property claims in scarce physical resources (document, vault, manufacturing plant) and contractual claims against parties who agreed to confidentiality. The combination is sufficient to keep the secret in practice over the long run.

The framework's predictions for the edge cases are clear. A burglar who breaks into the vault commits aggression against physical property and is liable on those grounds. An employee who leaks the recipe in violation of the NDA has accepted payment conditionally and committed fraud or breach of contract on those grounds. A chemist who buys a bottle off the shelf and reverse-engineers the formula commits no aggression: the chemist has used his own equipment and his own mind to arrive at a pattern that happens to match the firm's. He owes the firm nothing. An independent inventor who arrives at the same recipe without ever encountering the firm's product commits no aggression for the same reason. These predictions match what a reasonable libertarian intuition holds, and they match Rothbard's position on patent in *Man, Economy, and*

State: independent discovery is not theft, and reverse engineering of legitimately acquired physical property is not theft either.

The trade secret holder's position is therefore strong without requiring property in the pattern. The holder's protection comes from physical security of the storage media, contractual obligations of the people granted access, plus the labor and capital invested in keeping the secret close. What the holder does not have is a claim against the world. That absence is the whole content of the disagreement with patent law, and it is the position Rothbard and Kinsella converge on, the same position the framework this book follows extends.

These three mechanisms do not include property rights in information content itself, claims against people who independently discover the same information, rights to prevent others from using their own property in certain ways, or control over information after voluntary, unconditional disclosure. If you tell a stranger your secret with no confidentiality agreement, you have no property claim to prevent them from sharing it. You controlled disclosure (self-ownership); you could have kept silent. Having chosen to speak, you cannot claim property rights in the patterns now in their mind.

6.5 Implications for Privacy Analysis

What ``Privacy Violation'' Means

Given this framework, privacy violation is not someone knowing information you wish they did not know; that may be unfortunate but is not a violation. In the standard case, privacy violation means someone accessing information through aggression against person or property.

Examples of violation include forcibly extracting information through torture or coerced testimony, trespassing to obtain information by breaking into home or device, breaching confidentiality contract which constitutes theft of conditional payment, and fraud to obtain information through impersonation or deception.

Examples of non-violation include observing someone in public, receiving information voluntarily shared, independently discovering information, and learning information from someone who was told without confidentiality agreement.

Reputation Is Not Property

A related case the framework rules out is that reputation is not property. A reputation is what other people think about a person. It exists in their minds and on their media, not in any resource the person being talked about owns. Defamation law, like trademark and copyright, asserts a property-like interest in something non-scarce, in this case what other people believe and what audiences attend to. The framework treats false speech the same way it treats true speech: legitimate response is counter-speech, withdrawal of patronage, and social withdrawal. Where speech plays a causal role in actual aggression, a witness lying on the stand to send the defendant to prison, an order to commit assault, incitement of a crowd to lynch, the speaker is liable for the aggression his speech causes, not for damaging the victim's reputation as such. The same conclusion this chapter reached for patents and copyrights applies here: state enforcement of reputation rights against peaceful speech is aggression, not protection.[11]

The Role of Technology

Technology changes what is possible, not what is legitimate.

Encryption protects physical property (your devices) from search. This is an exercise of property rights, not a new right created by technology.

Anonymous networks protect self-ownership by enabling communication without revealing identity. This extends the natural privacy of in-person cash transactions to digital contexts.

Privacy technology implements existing rights. It does not create new rights or new categories of property.

Free and Open-Source Software as the Working Example

Free and open-source software[12] is the standing proof that the non-scarcity framework describes a functioning economy. Source code is a pattern; once released under a permissive or copyleft license, the pattern can be copied, modified, and redistributed by anyone with the physical media to hold it. No artificial-scarcity enforcement is invoked, and the software still gets written, maintained, and improved at a scale that now runs most of the world's servers, embedded devices, and cryptographic infrastructure. Linux,[13] OpenSSH, OpenSSL, GnuPG, Bitcoin Core, Tor,[14] Signal,[15] and the compiler toolchains that produce them are all instances of valuable software produced without the patent and copyright assumptions mainstream economics treats as necessary conditions for software production.

The revenue models that support this work are consistent with the framework. Developers are paid for scarce inputs: their time, their attention, their domain expertise, and their reputations for delivering working code. Firms pay for support contracts, custom development, hosted instances, and hardware that runs the software. Foundations collect voluntary contributions from users who benefit from the work and want it to continue. None of these arrangements requires property claims over the patterns themselves; they require ordinary contracts over the scarce labor and services that produce and maintain the patterns. The license attached to the code is a disclaimer of certain state-granted privileges, not an assertion of property in the content.

The licenses in wide use fit the same logic. Permissive licenses (MIT, BSD, Apache) allow any downstream use and rely on the underlying fact that the content is not property; they disclaim the state-granted copyright privilege entirely. Copyleft licenses (GPL, AGPL)[16] use the copyright privilege instrumentally: a conditional grant that requires downstream distributors to extend the same freedoms to their users, so that the artificial scarcity cannot be closed back up at a later stage in the supply chain. Both strategies treat copyright as a privilege to be neutralized, and both produce software with longer service lives, broader audit exposure, and greater

resistance to supplier capture than the closed-source alternatives the framework's critics predicted would outcompete it.

The practical implication for privacy is direct. Every privacy-critical primitive this book relies on, including the cryptographic libraries in Chapter 14, the anonymity networks in Chapter 17, the Bitcoin implementations in Chapters 18 through 20, the zero-knowledge tooling in Chapter 15, and the secure-computation runtimes in Chapter 16, is free software. Their auditability is what makes the resistance claims in Chapter 5 credible in the first place; a privacy tool whose source is closed is a privacy tool whose adversary-model includes its own author. Free software is the operating environment in which the book's technical arguments hold, and its existence is the economic vindication of the non-scarcity framework developed in this chapter.

``Surveillance Capitalism'' and Information Economics

Whether companies that collect data about users commit privacy violation depends on the terms.[17] If users agree to data collection (however unwisely) in exchange for services, no violation occurs: it is voluntary exchange. The exchange may be foolish, the terms may be buried in dense agreements, but consent was given.

If companies collect data without consent or through deception that exceeds the scope of agreements, this may violate contract or involve fraud.

But the data collected, once in the company's possession, is stored on their media. They own their servers. The information patterns on those servers are not "your property" that you can reclaim. The remedy for excessive data collection is not property claims but better contracts and privacy-preserving competition.[18]

Chapter Summary

Property rights apply to scarce resources, and scarcity is what creates the conflicts property rights exist to resolve. Information content is non-scarce: unlimited parties can hold the same idea without conflict. It cannot be property, and the mechanisms patents and

copyrights use to pretend otherwise create artificial scarcity through state violence, granting some parties control over how others may arrange their own physical media. That is restriction enforced by aggression, not protection of rights.

Privacy does not need information-as-property to be defensible. Self-ownership protects mind and body, physical property protects devices and papers and homes, and contract creates enforceable confidentiality obligations through voluntary agreement. These three mechanisms cover the person and their possessions, not abstract patterns of information. They do not reach independent discoverers, downstream use of other people's property, or information voluntarily disclosed without condition, and they should not; the framework rejects ownership claims over patterns and limits ownership to scarce things. Free and open-source software is the standing empirical confirmation of the framework. Source code is a non-scarce pattern released under licenses that disclaim or neutralize the copyright privilege, and the resulting software runs most of the world's cryptographic and communications infrastructure while developers are compensated for scarce inputs (time, attention, expertise) through ordinary contracts for labor, support, and services.

What this chapter sets is the criterion, not its full application. A complete theory of commercial data practices needs Chapter 11's treatment of deceptive collection, and the technical means of enforcing privacy when contract alone does not suffice belong to the later chapters on cryptography, anonymous networks, and private monetary layers. The privacy tools this book relies on implement existing rights, and the next three chapters carry the framework into exchange, capital theory, and monetary analysis.[19]

Notes

[1]Thomas Jefferson, letter to Isaac McPherson, August 13, 1813, in *The Writings of Thomas Jefferson*, ed. Andrew A. Lipscomb and Albert E. Bergh (Washington, DC: Thomas Jefferson Memorial Association, 1905), 13:326–338, available at the Library of Congress at `https://founders.archives.gov/documents/Jefferson/03-06-02-0322`. The letter is the canonical statement of the anti-intellectual-property position by

one of the officials who would, in his role as Secretary of State and then President, administer the American patent system. The modern restatement of Jefferson's argument is Stephan Kinsella, *Against Intellectual Property* (Auburn, AL: Ludwig von Mises Institute, 2008), 15, which this chapter develops in detail and which provides the contemporary theoretical framework.

[2]Kinsella's fullest statement is *Against Intellectual Property*, originally published as "Against Intellectual Property," *Journal of Libertarian Studies* 15, no. 2 (2001): 1–53. See also his blog at `https://stephankinsella.com` for ongoing elaboration and responses to critics.

[3]Rothbard's treatment in *Man, Economy, and State* (cited at Chapter 3, note 3), chapter 10 §7, distinguishes patent (a state-granted monopoly that suppresses independent inventors and is therefore illegitimate) from copyright reframed as a contractual restriction freely accepted by purchasers (which Rothbard argues is in principle legitimate as a private agreement, though he acknowledges it does not bind third parties who never agreed). The distinction was a real intellectual contribution and is the version of the libertarian-Austrian position most commonly cited in the older Mises Institute orbit. Eric Voskuil's libbitcoin "Patent Resistance Principle" (`https://github.com/libbitcoin/libbitcoin-system/wiki/Patent-Resistance-Principle`) works within the same Rothbardian register at the protocol-design level, treating patents as a state-granted monopoly that Bitcoin's open-source-license-and-prior-art posture is built to neutralize.

[4]Kinsella's developments since *Against Intellectual Property* (2001) sharpen the contract-route argument by showing what it does and does not deliver. A contract binds its parties; it does not bind third parties; downstream readers of a book never agreed to the publisher's terms; therefore "copyright as contract" cannot reach what statutory copyright reaches in operation (binding all third parties, including those who acquire the work without notice). Kinsella's title-transfer theory of contract, developed in *The Title-Transfer Theory of Contract* (Papinian Press Working Paper No. 1, 2024), explains why obligations transfer with consent and not otherwise. The disagreement between Rothbard's older treatment and Kinsella's later one is narrower than it appears: both reject patent, both accept that a contract binds its signatories on the conditions they agreed to, and both deny that the contract route delivers a claim against third parties. The book's privacy argument depends on the points the two positions share, not on the points where they diverge.

[5]On scarcity as the foundation of property, see Murray N. Rothbard, *The Ethics of Liberty* (cited at Chapter 2, note 3), chapters 6–8, and Hans-Hermann Hoppe, *A Theory of Socialism and Capitalism* (cited at Chapter 3, note 7), chapter 2. On the economics of non-rivalrous goods and the empirical case against intellectual monopoly, see Michele Boldrin and David K. Levine, *Against Intellectual Monopoly* (Cambridge: Cambridge University Press, 2008), `https://levine.sscnet.ucla.edu/general/intellectual/against.htm`.

[6]On the possession-vs-ownership distinction, see Stephan Kinsella, "Another Way to Explain the Problem with IP: Resources v. Knowledge, Ownership v.

Possession," *Center for the Study of Innovative Freedom* (February 2017), https://c4sif.org/2017/02/another-way-to-explain-the-problem-with-ip-resources-v-knowledge-ownership-v-possession/. The Roman-law sources are summarized in J. A. C. Thomas, *Textbook of Roman Law* (North-Holland, 1976), §§13–14. For the modern statement of the same distinction within Kinsella's framework, see *Legal Foundations of a Free Society* (Papinian Press, 2023), Appendix I, on possession, ownership, and the right to possess.

[7]On the distinction between property and privilege, see Rothbard, *The Ethics of Liberty* (cited at Chapter 2, note 3), chapter 10, and Lysander Spooner, "The Law of Intellectual Property" (1855), which argued for intellectual property but illuminates the issues.

[8]On intellectual property as a nonconsensual negative servitude, see Stephan Kinsella, "Intellectual Property Rights as Negative Servitudes," *Center for the Study of Innovative Freedom* (June 2011), https://c4sif.org/2011/06/intellectual-property-rights-as-negative-servitudes/. The framing is restated in Kinsella, *The Problem with Intellectual Property* (Papinian Press Working Paper No. 2, 2025), §III.A.2. On the distinction between consensual and nonconsensual servitudes in property law, see A. N. Yiannopoulos, *Louisiana Civil Law Treatise: Predial Servitudes*, 4th ed. (West, 2013), and the Roman-law treatment in Justinian, *Digest*, Book 8.

[9]The Lockean intuition that mixing one's labor with a resource gives one ownership of it is set out in John Locke, *Second Treatise of Government* (1689), §27 ("Whatsoever then he removes out of the State that Nature hath provided, and left it in, he hath mixed his Labour with, and joyned to it something that is his own, and thereby makes it his Property"); the canonical online edition is at https://www.gutenberg.org/ebooks/7370. On the rejection of creation as a source of property rights, see Stephan Kinsella, "Libertarian Lockean Creationism," *Center for the Study of Innovative Freedom* (May 2025), https://c4sif.org/2025/05/libertarian-lockean-creationism/, and Kinsella, *The Problem with Intellectual Property* (Papinian Press, 2025), §III.A.1. The same point is made at length in Kinsella, "Locke on IP; Mises, Rothbard, and Rand on Creation, Production, and 'Rearranging,"' Center for the Study of Innovative Freedom (2010). For the broader case that IP is not property at all, see Kinsella, "IP Is Not (Not) Property" (April 2025), https://c4sif.org/2025/04/ip-is-not-not-property/. Roderick T. Long, "Owning Ideas Means Owning People," *Cato Unbound* (November 19, 2008), http://www.cato-unbound.org/2008/11/19/roderick-long/owning-ideas-means-owning-people, and Long's earlier "The Libertarian Case Against Intellectual Property Rights," *Formulations* 3, no. 1 (Autumn 1995), http://freenation.org/a/f31l1.html, develop the implication that IP claims trace back to claims over other persons.

[10]On contract as the basis for confidentiality, see Randy Barnett, "A Consent Theory of Contract," *Columbia Law Review* 86 (1986): 269–321, and Kinsella's discussions of trade secrets in *Against Intellectual Property* (cited in note 2 above).

[11]On reputation as non-property and the incompatibility of defamation law with the framework, see Stephan Kinsella, "Defamation as a Type of Intellectual Property" (cited at Chapter 2, note 10), and the earlier post at `https://c4sif.org/2024/07/defamation-type-of-ip/`. The framework treats reputation as the contents of other people's beliefs, which are non-scarce and therefore non-ownable; the same scarcity-based reasoning that rules out patents and copyrights rules out reputation rights. Murray N. Rothbard's classic statement is "Knowledge, True and False," in *The Ethics of Liberty* (cited at Chapter 2, note 3), chapter 16, which argues that defamation actions are illegitimate because no one owns the contents of others' minds. Walter Block, "The Slanderer and Libeler," in *Defending the Undefendable* (Fleet Press, 1976), reaches the same conclusion through a different route. On the causal-incitement cases that are sometimes mistaken for defamation, see Kinsella, "Causation and Aggression," in *Legal Foundations of a Free Society* (Papinian Press, 2023), chapter 8.

[12]Richard Stallman, "What is Free Software?", GNU Project, `https://www.gnu.org/philosophy/free-sw.html`. The canonical definition of free software: software that grants users the freedoms to run, study, modify, and redistribute it. The Free Software Foundation, founded by Stallman in 1985, provides the ideological and legal infrastructure (including the GPL) behind the examples this section cites. For the broader movement, see Sam Williams, *Free as in Freedom: Richard Stallman's Crusade for Free Software* (O'Reilly Media, 2002), `https://www.oreilly.com/openbook/freedom/`. The term "open-source" was coined in 1998 by Christine Peterson and formalized by the Open Source Initiative; see `https://opensource.org/history`.

[13]Linus Torvalds announced the Linux kernel in a 1991 Usenet post to comp.os.minix: "I'm doing a (free) operating system (just a hobby, won't be big and professional like gnu) for 386(486) AT clones." The kernel's source is at `https://kernel.org`. Linux is the paradigm case of large-scale software production without intellectual-property enforcement; see Steven Weber, *The Success of Open Source* (Harvard University Press, 2004), for the political economy of its development model.

[14]The Tor Project, `https://www.torproject.org`. Tor (The Onion Router) is a free, open-source anonymity network that routes traffic through a series of volunteer-operated relays, encrypting each layer so no single relay knows both source and destination. Originally developed by the U.S. Naval Research Laboratory; see Roger Dingledine, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium* (2004), `https://svn.torproject.org/svn/projects/design-paper/tor-design.pdf`. Tor is the primary anonymity-network primitive this book's Chapter 17 relies on.

[15]Signal, `https://signal.org`. Signal is a free, open-source encrypted messaging application developed by the Signal Foundation (formerly Open Whisper Systems). Its cryptographic protocol — the Signal Protocol, combining the Double Ratchet Algorithm and the X3DH key agreement — is the basis for

end-to-end encryption in WhatsApp, Google Messages, and several other mainstream applications. See Moxie Marlinspike and Trevor Perrin, "The Double Ratchet Algorithm" (2016), `https://signal.org/docs/specifications/doubleratchet/`. Signal is cited in this book as an instance of privacy-critical infrastructure produced under open-source licensing.

[16]Richard Stallman, *GNU General Public License*, version 2 (1991) and version 3 (2007), `https://www.gnu.org/licenses/gpl-3.0.html`. The GPL is the founding copyleft instrument: it grants all the freedoms of free software while requiring that any distributed derivative work carry the same license, preventing downstream parties from closing the code. The Affero GPL (AGPL) extends this to network-deployed software. The legal mechanics are analyzed in Eben Moglen, "Enforcing the GNU GPL," *Columbia Law School* (2001), `https://moglen.law.columbia.edu/publications/lu-12.html`. The strategy described in the body text — using the state-granted copyright privilege instrumentally to prevent re-enclosure — is Stallman's invention and represents the clearest example of the framework in section 6.2: using an existing legal mechanism against the grain of its original purpose.

[17]Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (PublicAffairs, 2019). Zuboff's argument is that platform firms extract "behavioral surplus" — data beyond what is needed to provide the service — and use it to build predictive products sold to advertisers. The section heading uses her term; the analysis differs: where Zuboff argues for a new property right in behavioral data, the framework here denies that data patterns can be property and locates the remedy in contract and competition rather than data-ownership claims. The disagreement is fundamental and is the precise point at which the Austrian property-rights framework and the progressive data-rights framework diverge. Kinsella's most recent restatements of the framework: *The Problem with Intellectual Property* (Papinian Press Working Paper No. 2, 2025), `https://stephankinsella.com/wp-content/uploads/publications/kinsella-problem-intellectual-property-v1.1-2025.pdf`, is the current book-chapter treatment, with *The Title-Transfer Theory of Contract* (Papinian Press Working Paper No. 1, 2024) as the companion contract paper. The libertarian-Austrian case at book length: Stephan Kinsella, *Against Intellectual Property* (Mises Institute, 2008), `http://www.stephankinsella.com/wp-content/uploads/publications/kinsella-against-IP-book.pdf`; his *You Can't Own Ideas: Essays on Intellectual Property* (Papinian Press, 2023) collects the more recent essays. The shorter pieces directly relevant to this chapter's framing: Kinsella, "Intellectual Property Rights as Negative Servitudes" (cited in note 8 above), develops the negative-servitude framing this chapter relies on; Kinsella, "Another Way to Explain the Problem with IP" (cited in note 6 above), develops the possession-vs-ownership distinction; Kinsella, "Libertarian Lockean Creationism" (cited in note 9 above), names the recurring error this chapter rejects. Roderick T. Long, "Owning Ideas Means Owning People" (cited in note 9

above), traces the IP claim to its implication that the holder owns part of other persons. On the mainstream-economics critique of intellectual monopoly, Michele Boldrin and David K. Levine, *Against Intellectual Monopoly* (cited in note 5 above), is the empirical treatment. Fritz Machlup, *An Economic Review of the Patent System*, Study No. 15 (U.S. Senate Subcommittee on Patents, 1958), is the classic institutional-economics study that concluded the patent system could not be justified on the evidence available. For historical context, Adrian Johns, *Piracy: The Intellectual Property Wars from Gutenberg to Gates* (University of Chicago Press, 2009), is the standard history. On copyright specifically, Lawrence Lessig, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity* (Penguin, 2004), is the canonical mainstream-progressive critique; Tom W. Bell, *Intellectual Privilege: Copyright, Common Law, and the Common Good* (Mercatus Center, 2014), is the more libertarian version. The opposing view, for fairness: William M. Landes and Richard A. Posner, *The Economic Structure of Intellectual Property Law* (Harvard University Press, 2003), is the standard law-and-economics defense of the existing system.

[18]On data and privacy in the digital age from a property-rights perspective, see Stephan Kinsella's ongoing writing at `https://stephankinsella.com` (in particular his essays tagged "intellectual property" and "Against IP Reader") and the Center for the Study of Innovative Freedom at `https://c4sif.org`. See also Kinsella, *You Can't Own Ideas: Essays on Intellectual Property* (Papinian Press, 2023), which collects his most recent treatments. On the "data ownership" framing specifically, see Jaron Lanier, *Who Owns the Future?* (Simon & Schuster, 2013), which argues for data as property from a non-libertarian standpoint; Kinsella's response is in his blog post "Data Ownership Misses the Point."

[19]Further reading on intellectual property from multiple perspectives.

7

EXCHANGE THEORY AND PRIVACY

> "Exchange is not an end in itself. It is the means by which men advance their welfare."
>
> Carl Menger, *Principles of Economics* (1871)[1]

Exchange is the foundation of social cooperation.[2] When individuals trade, they signal valuations, coordinate production, and create wealth impossible through isolated action. Price signals emerging from exchange[3] enable the economic calculation that makes complex societies possible.

People exchange under surveillance constantly, and markets function, however imperfectly, in surveilled environments. The claim here is narrower: privacy enhances exchange. Surveillance distorts market processes in identifiable ways, and privacy protection improves many exchanges while preserving some marginal exchanges that surveillance would otherwise chill away.

7.1 The Logic of Exchange

Exchange as Mutual Benefit

Exchange occurs when parties expect to benefit. Alice values what Bob has more than what she offers; Bob values what Alice offers more than what he has. Both expect to be better off after trading than before.

Subjective value operates here directly. No objective measure exists making the exchange "fair" or "equal." Each party evaluates from their own perspective, according to their own preferences and circumstances. If both prefer to trade, both gain.

Exchange is thus positive-sum. Unlike theft or redistribution, where one party's gain is another's loss, voluntary exchange creates value for all participants. The total wealth of society increases through trade.

Exchange Requires Information

For exchange to occur, parties need information. They require knowledge of opportunity: they must know exchange is possible, find each other, communicate, and identify potential trades. They also require knowledge of terms: they must understand what is being offered and requested, since misunderstanding terms produces regret instead of mutual benefit. Finally, they require knowledge sufficient for evaluation: each party must have enough information to determine whether the exchange serves their interests.

But exchange does not require complete information. Parties routinely trade with imperfect knowledge of product quality and counterparty reliability, all under uncertain future conditions. Uncertainty is inherent in action; exchange operates within uncertainty, not by eliminating it.

Exchange Requires Deliberation

Before agreeing to trade, each party deliberates. They consider what they are giving up and what they are getting, whether this trade is better than alternatives, what the risks are and what could go wrong, and whether the trade serves their goals.

This deliberation, as Chapter 3 established, is internal.[4] It occurs in the mind of the acting individual. The conclusions, the final valuations and choices, depend on subjective factors inaccessible to external observers.[5]

7.2 How Privacy Enhances Exchange

Protected Deliberation

Deliberation works best when protected from external interference.

If Alice knows her thinking is being monitored, her deliberation may change.[6] She may consider how her thoughts will be perceived, avoid conclusions that might draw criticism, shape her reasoning to satisfy observers, or second-guess herself based on expected reactions.

Such thinking is performance shaped by observation, not deliberation serving Alice's interests. The "choices" emerging from monitored deliberation may not reflect Alice's actual preferences.

Privacy protects deliberation by creating space for authentic evaluation. When Alice's thinking is private, she can evaluate options according to her own standards without concern for observer reactions. Her conclusions are more likely to reflect her actual interests.

For exchange, this means: private deliberation produces better-informed trading decisions. As Chapter 3 established, deliberation is inherently internal to the actor; external observation cannot access the subjective valuation process but can distort it. Parties who can think freely evaluate opportunities more accurately than parties constrained by observation.

Negotiation Without Exposure

Negotiation is strategic interaction.[7] Each party tries to achieve favorable terms while reaching agreement.

Effective negotiation requires controlled disclosure, because every piece of information a party reveals shifts the bargain.[8] Revealing your maximum willingness to pay weakens your bargaining position, exposing urgency invites exploitation, showing your alternatives signals your walkaway point, and disclosing future plans enables strategic positioning by counterparties.

If all negotiation information were transparent, bargaining would tilt sharply toward the side with more patience and better

alternatives, or less immediate need. Strategic interaction requires strategic information control.

Privacy enables negotiation by protecting information parties need to control. Each side can reveal what serves their interests while concealing what would weaken their position. The asymmetric disclosure is not dishonesty but appropriate boundary management: strategic interaction assumes parties manage what they reveal, and effective negotiation depends on the ability to do so.

Confidential Terms

Many exchanges benefit from confidential terms. Price confidentiality allows sellers to offer different prices to different buyers based on circumstances; if all prices were public, this flexibility would disappear and could prevent mutually beneficial trades. Custom arrangements allow terms to be tailored to specific situations, whereas public exposure would pressure parties toward standardized terms even when customization serves both parties. Revealing contract terms may inform competitors, who can then undercut or copy arrangements that required investment to develop.

Privacy enables parties to structure exchanges according to their specific needs without exposing arrangements to competitive copying or third-party interference.

Trust Building Over Time

Long-term exchange relationships require trust. Trust develops through repeated interaction and consistent performance, then deepens through graduated disclosure and mutual investment in relationship-specific assets.

Privacy supports trust building by enabling graduated disclosure. Parties can reveal more as trust increases without being forced into premature transparency. The relationship develops at its own pace, not forced by external observation.

Surveillance disrupts trust building by removing control over disclosure pace. If all interactions are observed, parties cannot manage the gradual revelation that natural trust development requires.

7.3 How Surveillance Distorts Exchange

The Chilling Effect

Surveillance chills exchange by introducing risks beyond the transaction itself.

If transactions are monitored, parties are likely to consider how the exchange will be perceived, whether the transaction could be used against them later, what inferences observers will draw, and whether the exchange is safe given who might be watching.

These considerations have nothing to do with whether the exchange benefits both parties. They are external factors imposed by surveillance that distort decision-making.

The effect, at the margin, is that some exchanges that would benefit both parties do not occur because of surveillance risk. Value that would have been created is not created. Market coordination is impaired.

Price Signal Degradation

Prices coordinate economic activity by communicating information about relative scarcity and value. Accurate prices depend on authentic exchange reflecting actual valuations.

Surveillance degrades price signals by chilling transactions that would occur without surveillance, biasing transactions toward surveilled-acceptable patterns, introducing strategic behavior to manage surveillance records, and reducing market participation by surveillance-averse parties.[9]

Prices emerging from surveilled markets reflect not just supply and demand but also surveillance avoidance. They are systematically distorted as information signals.

Strategic Behavior Shift

Under surveillance, parties shift from serving their interests to managing their records.

Instead of asking "What exchange serves my goals?", parties may ask "What exchange looks appropriate to observers?" Decision cri-

teria are likely to shift from authentic preference toward appearance management.

Where the shift occurs, the effect is economically destructive. Resources flow not to their highest-valued uses but to their most surveillance-acceptable uses. The allocation is distorted by external judgment, not guided by participant valuations.

Third-Party Interference

Surveillance enables third-party interference with exchange.

If transactions are monitored, parties with access to monitoring data can intervene in transactions they disapprove of, tax transactions they can observe, regulate exchanges based on observed patterns, and target participants for political or competitive reasons.

This interference is possible only because surveillance provides the information enabling it. Privacy forecloses interference by denying the information it requires.

Regime Uncertainty and the Investment Chill

Robert Higgs's analysis of the Great Depression and the postwar recovery named a condition this chapter's argument depends on.[10] Regime uncertainty is the condition in which private actors cannot predict whether the state will change the legal rules governing property, contract, regulation, and taxation. Under regime uncertainty, the expected return on long-horizon investment is discounted by the probability that the rules under which the investment pays off will be altered before the payoff arrives. Capital formation shifts toward shorter horizons, and projects whose value depends on a stable legal regime are delayed or abandoned.

Surveillance amplifies regime uncertainty along a specific axis. A transaction recorded today is a transaction that can be reinterpreted and prosecuted under whatever rules are in force at any later date. An entrepreneur who enters a line of business whose legal status is ambiguous must discount the expected return by the probability that the rules will tighten and that the records of the ambiguous

period will be used to construct a case. The less private the record, the larger the discount.

The effect compounds when the categories of permissible activity shift with political weather. Speech protected in one administration invites investigation in the next. Financial flows permitted in one regulatory era become grounds for de-banking in another. The individual contemplating long-term commitment to any activity whose legal status depends on interpretation has been handed a legibility cost that the investment must overcome before it can produce a return. Privacy at the point of transaction narrows the variance of future regulatory exposure to the rules currently stated, and this is the closest substitute for stable law that a private actor can build without state cooperation.

7.4 Exchange Can Occur Under Surveillance

The claim is not that surveillance makes exchange impossible. Exchange occurs constantly under surveillance. But it is distorted.

Existing Surveilled Exchange

Most modern exchange is surveilled to some degree. Financial transactions are monitored and reported. Online purchases create data trails, while communications are subject to interception and physical movement is tracked through various means.

Markets still function in this environment. Prices emerge, goods exchange hands, and broader economic coordination still occurs.

The Distortions Are Real But Limited

The distortions described above are real. But their magnitude depends on how extensive the surveillance is, how much parties care about being observed, how likely intervention based on surveillance is, and what alternatives to surveilled exchange exist.

When surveillance is light and consequences are unlikely, distortions may be small, especially where no realistic alternatives exist. Parties then accept surveillance costs as part of doing business.

When surveillance is heavy and consequences are likely, distortions are larger, especially where alternatives exist. Parties may shift to unsurveilled channels or else forgo exchange entirely.

The Marginal Cases

Even if most exchange continues under surveillance, the marginal cases still shape outcomes:

The exchanges that do not occur are lost value. Trades that would benefit both parties are prevented by surveillance risk.

The exchanges that are distorted misallocate resources. Decisions shaped by surveillance management instead of actual preference produce inferior outcomes.

The exchanges that shift to alternatives impose adaptation costs. Resources spent creating and maintaining privacy tools are resources not available for other purposes.

Privacy protection does not make exchange possible in the abstract. It improves exchange in general and preserves some exchanges that surveillance would otherwise chill: more transactions at the margin, less distortion, more accurate prices, more efficient allocation.

7.5 Specific Exchange Contexts

Employment

Employment is ongoing exchange: labor for compensation. The relationship involves repeated negotiation and performance evaluation, along with mutual assessment that unfolds over months and years.

Privacy affects employment at every stage. A job seeker whose current employer discovers an active search may be preemptively terminated or cut off from further development. Salary negotiation requires concealing reservation wages; a candidate who reveals what they would accept anchors the offer at that floor, well below what the position would otherwise yield. Within ongoing employment, constant monitoring shifts employees from substan-

tive work toward performance of work, optimizing for observable metrics instead of actual contribution. Meta-analyses find no significant relationship between electronic monitoring and performance but positive relationships with stress and counterproductive work behavior.[11] Creative work, complex problem-solving, and discretionary effort all diminish under observation; the employer who monitors completely may see everything the employee does while ensuring that what the employee does has less value.

The asymmetry compounds the problem. Employers who can monitor while employees cannot observe employer intentions gain systematic advantage in wage negotiation, performance evaluation, and termination. The employee negotiating a raise cannot know whether the employer has already decided to eliminate the position; the worker asked to take on additional responsibilities cannot know whether promotion or exploitation awaits. An employee planning to leave needs confidentiality to maintain their current position while seeking alternatives, because discovery typically accelerates termination, wipes out final bonuses, and poisons references.

Professional Services

Professional services depend on confidentiality and legal systems recognize this through privilege doctrine. Attorneys cannot be compelled to reveal client communications. Physicians maintain patient confidentiality. These professions cannot function without protected disclosure.[12]

The structure is the same across the privileged professions. A client seeking legal advice must describe the facts creating the liability; a patient must describe symptoms and behavior; an investor must reveal their financial position and risk tolerance. Without protection, clients withhold relevant information, professionals advise on incomplete understanding, and the service collapses. Professional privilege was never a gift to professionals. It is institutional recognition that certain exchanges cannot function without privacy.

Business-to-Business

Business exchange involves proprietary information more complex than consumer transactions. Suppliers learn customer demand forecasts, customers learn supplier cost structures, and partners learn strategic plans. Each disclosure creates competitive vulnerability if the information reaches competitors.

Businesses handle this through graduated disclosure calibrated to relationship depth. Initial transactions reveal only what is necessary for that specific exchange. As relationships deepen and trust develops, parties share more strategic information enabling closer coordination. This graduated pattern, paralleling personal trust development, is impossible without privacy that allows parties to control what they reveal.

Price negotiation illustrates the stakes. A supplier who knows that a customer has no alternatives can extract higher prices, and a customer who knows that a supplier is desperate for revenue can demand concessions; both parties therefore conceal information about their alternatives and urgency, along with the constraints they face. Transparent negotiation would collapse into exploitation of the weaker side.

Joint ventures and partnerships require sharing information that could be exploited if the partnership fails. Development plans and customer lists, along with technical capabilities shared with partners, become competitive weapons if those partners become rivals. The exchange of information is itself the thing being traded. Without privacy protecting shared information from third parties and constraining use if the partnership ends, businesses would refuse to share what productive cooperation requires.

Supply chain relationships demonstrate how privacy enables exchange across organizational boundaries. A manufacturer sharing demand forecasts with suppliers enables those suppliers to optimize production, benefiting both parties. But those forecasts also reveal the manufacturer's expectations about their own market. Privacy allows functional disclosure to supply chain partners without strategic disclosure to competitors.

Consumer Markets

Consumer exchange appears simpler but involves privacy interests that accumulate across transactions. Any single purchase reveals little. Aggregate purchase data reveals preferences, circumstances, health conditions, political views, relationships, and vulnerabilities.

Consider how purchase patterns function as surveillance. Pharmaceutical purchases reveal health conditions; book purchases reveal intellectual interests and political leanings; grocery purchases reveal dietary restrictions that may indicate religious practice or health status. Location patterns reveal workplace and residence, along with broader associations. Financial data reveals economic status and cash flow, along with payment reliability. None of these disclosures is necessary for the transaction itself. Each is surveillance surplus extracted from the exchange.

The privacy tools that traditionally protected consumer exchange are disappearing. Cash enables purchase without identification, but cash acceptance declines as payment infrastructure shifts digital.[13] Anonymous in-store purchase requires physical presence, but retail shifts online where every transaction is logged. Consumer privacy once required no affirmative action; now it requires deliberate tool adoption against default surveillance.

The aggregate dimension distinguishes consumer privacy from the other contexts. An employer surveilling a specific employee has a specific purpose. Corporate consumer surveillance is complete: transaction histories and location traces, along with clickstreams, are accumulated across all consumers to enable targeting, manipulation, and prediction.[14] The consumer's privacy interest is not in any single transaction but in resisting the aggregation that turns innocuous purchases into complete profiles.

This creates collective action problems individual privacy tools cannot solve. If one consumer uses cash while others use tracked payment, the cash user maintains privacy but cannot prevent the profile construction that aggregate data enables. Consumer privacy requires either mass adoption of privacy tools or structural changes

that prevent aggregation. Both face coordination challenges that employment or professional privacy does not.

Chapter Summary

Exchange is mutual benefit through trade. It requires information, deliberation, and agreement, and through price signals it coordinates production and consumption across the economy. Privacy enhances exchange at every layer: protected deliberation produces better-informed decisions, controlled disclosure lets negotiation work without collapsing into exploitation of the weaker side, confidential terms let parties structure arrangements their competitors cannot copy, and graduated revelation lets trust develop at its own pace. Surveillance distorts the same processes. Monitored deliberation shifts decisions toward appearance management, chilled transactions do not occur, and third parties intervene in exchanges they can observe.[15][16]

The claim is comparative, not absolute. Exchange occurs under surveillance constantly; privacy does not make exchange possible but makes it better, preserving marginal transactions and reducing systematic distortion. The economic case stands independently of the normative one. Even without Chapter 4's ethical argument, the coordination gains privacy provides would justify protecting it. The axiomatic foundation in Chapter 3, the ethical prohibition in Chapter 4, and the tool-level implementation in Parts V and VI complete a picture this chapter treats strictly in economic terms.

Notes

[1]Carl Menger, *Principles of Economics*, trans. James Dingwall and Bert F. Hoselitz (Auburn, AL: Ludwig von Mises Institute, 2007; German original 1871), chapter 4 ("The Theory of Exchange"), esp. pp. 175–191. The founding text of the Austrian School, in which Menger derives the exchange relationship from marginal utility without recourse to labor theory or equilibrium price. Mises develops the same relationship more fully in *Human Action* (cited at Chapter 3, note 1), 195 and part IV generally, which this chapter draws on throughout.

[2]On exchange as the foundation of social cooperation, see Mises, *Human Action* (cited at Chapter 3, note 1), Part II, "Action Within the Framework of Society," and

Murray N. Rothbard, *Man, Economy, and State* (cited at Chapter 3, note 3), chapters 2–3.

[3]Israel M. Kirzner, *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973). Kirzner's central contribution is the concept of entrepreneurial alertness: profit opportunities exist in the market because prices are not yet equilibrated, and entrepreneurs who notice these gaps drive the price-coordination process. The "price signals emerging from exchange" sentence in the introduction draws on this: it is the entrepreneurial function, not a mechanical equilibrating machine, that converts dispersed valuations into useful price information. See also Kirzner, *The Meaning of Market Process* (Routledge, 1992), for the fuller statement of his market-process theory, and *Discovery and the Capitalist Process* (University of Chicago Press, 1985).

[4]On deliberation and action, see Mises, *Human Action* (cited at Chapter 3, note 1), chapter 1, "Acting Man," and the analysis in Chapter 3 of this book.

[5]On subjective value theory and exchange, see Carl Menger, *Principles of Economics* (cited at Chapter 3, note 5), 175–225 (the exchange chapter, distinct from the subjective-value chapter cited at Chapter 3, note 5).

[6]On surveillance effects on behavior, see Michel Foucault, *Discipline and Punish: The Birth of the Prison*, trans. Alan Sheridan (New York: Vintage, 1995; French original 1975). The panopticon analysis is the standard reference for the behavioral effect of perceived observation; the economic analysis here is Austrian, not Foucauldian.

[7]Thomas C. Schelling, *The Strategy of Conflict* (Cambridge: Harvard University Press, 1960). Schelling's game-theoretic analysis of bargaining, commitment, and focal points is the standard reference for strategic interaction in negotiation. His central insight — that parties in negotiation often benefit from limiting their own options, communicating credible commitments, and managing information asymmetries — grounds the argument in this section that privacy enables negotiation by letting each party control what it reveals. Schelling received the Nobel Memorial Prize in Economic Sciences in 2005. For applications to information asymmetry in bargaining specifically, see also Roger B. Myerson and Mark A. Satterthwaite, "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 29 (1983): 265–281.

[8]Friedrich A. Hayek, "The Use of Knowledge in Society," *American Economic Review* 35, no. 4 (1945): 519–530, `https://www.jstor.org/stable/1809376`. The canonical paper on prices as decentralized knowledge-coordination signals; the negotiation argument here extends Hayek's point that the relevant information is dispersed and cannot be aggregated by an outside observer without distorting the bargain. See also Israel M. Kirzner, *Market Theory and the Price System* (Indianapolis: Liberty Fund, 2011; original Van Nostrand, 1963), on the entrepreneurial function inside the price-coordinated market process.

[9]On price signals and economic calculation, see Mises, *Human Action* (cited at Chapter 3, note 1), Part III, "Economic Calculation," and Mises, *Socialism: An Economic and Sociological Analysis*, trans. J. Kahane (Indianapolis: Liberty Fund,

1981; German original 1922), Part II. The calculation argument names the specific function prices perform that surveillance distortion impairs.

[10]Robert Higgs, "Regime Uncertainty: Why the Great Depression Lasted So Long and Why Prosperity Resumed after the War," *The Independent Review* 1, no. 4 (Spring 1997): 561–590. The argument is extended in Robert Higgs, *Depression, War, and Cold War: Studies in Political Economy* (New York: Oxford University Press, 2006), and in "Regime Uncertainty: Pirrong Responds to Krugman," *The Independent Review* 16, no. 2 (Fall 2011): 243–246. Higgs is not Austrian in the strict methodological sense, but his concept is Austrian-compatible and has been taken up in Austrian political-economy writing since it was introduced. The extension to surveillance-induced uncertainty made here is this chapter's.

[11]Daniel M. Ravid et al., "The Impact of Electronic Monitoring on Employees' Job Satisfaction, Stress, Performance, and Counterproductive Work Behavior: A Meta-Analysis," *Computers in Human Behavior Reports* 8 (2022). See also Paul Glavin and Scott Schieman, with Alex Bierman, "Private Eyes, They See Your Every Move: Workplace Surveillance and Worker Well-Being," *Socius* 10 (2024).

[12]On attorney-client privilege, the canonical modern statement is *Upjohn Co. v. United States*, 449 U.S. 383 (1981), and the leading treatise treatment is John Henry Wigmore, *Evidence in Trials at Common Law*, ed. John T. McNaughton, vol. 8 (Boston: Little, Brown, 1961), §§2290–2329. On the federal psychotherapist-patient privilege, *Jaffee v. Redmond*, 518 U.S. 1 (1996). On physician-patient confidentiality as professional ethics, the Hippocratic Oath is the founding statement; the contemporary American codification is the AMA *Code of Medical Ethics*, Opinion 3.2.1 ("Confidentiality"). The economic-functional account given in this chapter is foreshadowed in Daniel J. Solove, *Understanding Privacy* (Harvard University Press, 2008), and in the contextual-integrity framework of Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford University Press, 2010); both are cited at Chapter 1, note 3.

[13]On the empirical decline of cash use, see the Federal Reserve Bank of Atlanta and Federal Reserve Bank of San Francisco, *Diary of Consumer Payment Choice* (annual series, 2012–present), which documents the falling cash share of consumer payments by transaction count and value. The Federal Reserve Bank of San Francisco's *2024 Findings from the Diary of Consumer Payment Choice* reports cash at 16% of payments by number, down from 31% in 2016. For the comparative international picture, see Bank for International Settlements, *Committee on Payments and Market Infrastructures: Red Book Statistics* (annual). On the policy and institutional drivers of digital-payment substitution, Brett Scott, *Cloudmoney: Cash, Cards, Crypto, and the War for Our Wallets* (Harper Business, 2022), is the most accessible book-length treatment.

[14]Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019). Zuboff documents how platform firms accumulate behavioral data across all consumer interactions to build predictive profiles sold to advertisers and used for behavioral modification. Her term "surveillance capitalism" names the business model; the

sentence this note anchors describes its mechanism: the aggregation of individually innocuous transactions into complete behavioral profiles. The framework of this chapter differs from Zuboff's in diagnosis and remedy — she argues for new data-property rights, this chapter locates the remedy in contract enforcement and privacy-preserving alternatives — but her empirical description of the aggregation dynamic is the reference point.

[15]On surveillance as a distortion of market processes, see Shoshana Zuboff, *In the Age of the Smart Machine: The Future of Work and Power* (Basic Books, 1988), the foundational study of how workplace monitoring shifts employee behavior from substantive performance to performance of performance. The same mechanism operating at the level of consumer markets is the subject of Zuboff's later *The Age of Surveillance Capitalism* (cited in note 15 above). On chilling effects in economic behavior under legal uncertainty, see Richard A. Posner, *Economic Analysis of Law*, 9th ed. (Wolters Kluwer, 2014), §§1.1–1.2, on the general deterrence framework; and Higgs, "Regime Uncertainty" (cited in note 8 above), on the investment-horizon compression that surveillance of transactions produces.

[16]Further reading on exchange, price signals, and market process. The foundational Austrian price-theory text is Mises, *Human Action* (cited at Chapter 3, note 1), parts III–IV, which treats calculation, prices, and the market as an integrated process. Murray N. Rothbard, *Man, Economy, and State* (cited at Chapter 3, note 3), chapters 2–7, presents the same material more systematically. The nineteenth-century classical-liberal parallel is Frédéric Bastiat, *Economic Harmonies*, trans. W. Hayden Boyers (Foundation for Economic Education, 1964; French original 1850), which develops the argument that exchange itself constitutes society; Bastiat's *Economic Sophisms* (1845) is the shorter companion. On the knowledge-coordination function of prices specifically, Hayek, "The Use of Knowledge in Society" (cited in note 6 above), is the canonical paper; Hayek's "Competition as a Discovery Procedure," in *New Studies in Philosophy, Politics, Economics, and the History of Ideas* (University of Chicago Press, 1978), extends the argument. Israel Kirzner, *Competition and Entrepreneurship* (University of Chicago Press, 1973), and *Discovery, Capitalism, and Distributive Justice* (Blackwell, 1989), develop the entrepreneurship side of the same framework. For the mainstream treatment, Alfred Marshall, *Principles of Economics*, 8th ed. (Macmillan, 1920), remains the neoclassical reference; George Stigler, "The Economics of Information," *Journal of Political Economy* 69, no. 3 (1961): 213–225, is the mainstream starting point for information economics. For the humanistic counterweight to caricatures of market theory, Wilhelm Röpke, *A Humane Economy: The Social Framework of the Free Market* (Henry Regnery, 1960), develops the German ordoliberal tradition. On the behavioral literature that connects surveillance to action, Shoshana Zuboff, *In the Age of the Smart Machine* (Basic Books, 1988), is the earlier, pre-*Surveillance Capitalism* treatment and is specifically about workplace monitoring. On privacy as a condition of authentic preference expression, Alan Westin, *Privacy and Freedom* (cited at Chapter 1, note 1), is the classic mid-twentieth-century study; Anita L. Allen's work, cited at Chapter 1, note 3, extends the analysis.

8

CAPITAL THEORY AND ENTREPRENEURSHIP

"Capital is produced means of production."

Eugen von Böhm-Bawerk, *The Positive Theory of Capital* (1889)[1]

"The entrepreneur is the driving force of the market economy."

Israel M. Kirzner, *Competition and Entrepreneurship* (1973)[2]

Privacy infrastructure is capital. It requires present sacrifice for future capability. Investment in encrypted communication systems and secure development environments, along with anonymous networks,[3] yields forms of coordination that direct approaches cannot achieve.

This chapter applies Austrian capital theory and entrepreneurship analysis to privacy. Böhm-Bawerk's insights about roundabout production explain why indirect methods of achieving privacy (building infrastructure) outperform direct methods (trusting institutions). Kirzner's entrepreneurial alertness explains how privacy innovations emerge through market discovery.

The analysis also illuminates a practical tradeoff: privacy often requires accepting inconvenience now for protection later. Time preference theory explains why some individuals invest in privacy infrastructure while others accept surveillance for immediate conve-

nience. Neither choice is irrational; they reflect different preferences about present versus future satisfaction.

Consider a merchant who wants to sell without building a permanent dossier about customers. The direct route is to use ordinary banking and a mainstream storefront while trusting intermediaries to behave, while the roundabout route is slower and asks the merchant to learn Bitcoin,[4] shift communication onto encrypted channels, build trade on rails that do not depend on platform surveillance. The second route demands present effort. It also creates future capabilities the first route cannot provide.

8.1 Privacy Infrastructure as Capital Goods

Capital Goods vs. Consumer Goods

Praxeology distinguishes capital goods from consumer goods by their relationship to human wants. Consumer goods satisfy wants directly, as food satisfies hunger; capital goods satisfy wants indirectly by enhancing future production, as a factory produces goods that satisfy wants. A single encrypted message or one-time anonymous purchase is a privacy consumer good, satisfying a want directly. Encrypted communication systems, secure development environments, and anonymous-network infrastructure are privacy capital goods: they enable future production of privacy benefits across many uses.

Higher-Order and Lower-Order Goods

Menger classified goods by their distance from final consumption:[5]

First-order goods (consumer goods) directly satisfy wants. Second-order goods produce first-order goods. Third-order goods produce second-order goods, and so on, with each order enabling production at the next lower level.

Privacy infrastructure exhibits this hierarchical structure. Cryptographic research and protocol development sit near the high end, user applications and privacy services at the low end, with develop-

ment frameworks and communication protocols in between; each higher order enables production at the next lower level.[6]

Investment in higher-order goods yields greater returns but requires longer time horizons and greater patience. Cryptographic research may take years before producing consumer-facing applications.[7]

Roundabout Production

Böhm-Bawerk showed that indirect (roundabout) production methods often yield superior results.[8] The direct approach uses simple tools for immediate results; the roundabout approach invests time building better tools, slower to start but yielding superior long-term capability.

Direct privacy relies on institutional promises and legal frameworks backed by trust: quick to implement but vulnerable to authority changes and institutional compromise. Roundabout privacy builds cryptographic infrastructure that enables mathematical verification, slower to develop but protected independently of institutional cooperation. The roundabout approach requires lower time preference, and the payoff is privacy that does not depend on others' cooperation.

8.2 Time Preference and the Privacy-Convenience Tradeoff

The Tradeoff

Privacy often requires inconvenience. Encrypted communication requires key management and anonymous transactions require additional steps; secure systems in general impose learning curves before they become usable.

This creates a real tradeoff: accept surveillance for immediate convenience, or accept inconvenience for future privacy protection.

For the merchant above, the inconvenience is concrete: extra setup and more skills to learn before the first private sale feels

ordinary. The later payoff is also concrete: cleaner settlement and fewer chokepoints watching the customer relationship.

Neither choice is inherently irrational. They reflect different time preferences, though time preference is not the only factor at work. Knowledge matters: users who do not understand surveillance risks cannot rationally weigh them. Convenience technologies often obscure their costs, making present benefits salient while hiding future risks. Social norms influence choices: if everyone uses surveilled services, privacy-protecting alternatives may lack network effects. Technical skill affects the cost of privacy: for users who find encryption easy, the tradeoff tilts toward privacy; for those who struggle, convenience weighs more heavily. Time preference remains central, but it interacts with information, skill, and social context.

Time Preference Theory

Time preference is the preference for present satisfaction over future satisfaction. All else equal, people prefer goods now to goods later. A dollar today is worth more than a dollar next year.

But time preference varies among individuals:

Individuals with high time preference have a strong preference for present satisfaction and are willing to accept future costs for present benefits. Individuals with low time preference have a weaker preference for present satisfaction and are willing to accept present costs for future benefits.

Application to Privacy

High time preference individuals may rationally choose surveillance-enabled services. The convenience is immediate and certain, while the privacy costs are future and uncertain. Discounting future costs heavily, the trade favors convenience.

Low time preference individuals may rationally choose privacy infrastructure. The inconvenience is temporary as learning occurs, while the privacy benefits compound over time. Valuing future benefits highly, the trade favors privacy.

This explains market segmentation in privacy technology. Some users adopt privacy tools despite inconvenience; others prefer convenient surveilled alternatives. Both are responding rationally to their own time preferences.

Infrastructure Investment as Low Time Preference

Building privacy infrastructure requires even lower time preference than using privacy tools. Developers invest years before seeing adoption. Protocol designers work without immediate compensation. Open source contributors sacrifice present income for future impact.[9]

This explains why privacy infrastructure development concentrates among individuals with unusually low time preference, often those with strong ideological commitment or unusual patience.

Market Coordination of Time Preferences

Markets coordinate different time preferences:

Low time preference individuals build infrastructure. They accept present sacrifice for future capability, investing in tools, protocols, and systems.

High time preference individuals use infrastructure. They benefit from others' investment without sharing the development burden.

The division looks like free-riding, but it is closer to specialization. Developers are compensated through eventual adoption and reputation, users pay through purchase prices or donations, and markets coordinate the different time preferences of each side through voluntary exchange.

8.3 Entrepreneurial Discovery in Privacy

Kirznerian Alertness

Israel Kirzner explained entrepreneurship as alertness to profit opportunities others miss:[10]

Entrepreneurs do not possess superior knowledge of given data. They notice opportunities that exist in the data but that others overlook. This alertness enables them to arbitrage between current market conditions and unrecognized possibilities.

Privacy technology exhibits Kirznerian discovery. Entrepreneurs notice coordination problems. Existing tools fail to protect certain communications, and market participants cannot coordinate without surveillance. Institutional solutions leave gaps that technology can fill.

They then develop solutions. New protocols address unmet needs and applications serve markets that incumbents ignored because surveillance was the business model.

Entrepreneurs profit from that alertness. Early adoption captures market position, and infrastructure investment creates durable competitive advantages.

Discovery of Privacy Needs

Privacy needs are not always obvious. Many people do not recognize their privacy is compromised until harm occurs. Entrepreneurs notice these unrecognized needs:

Some needs involve latent demand: people would value privacy if they understood what they were losing, and entrepreneurs educate while serving. Other needs are unarticulated: people experience identity theft or embarrassing leaks, along with chilling effects, without connecting them to privacy, and entrepreneurs make the connection. Still other needs are future-oriented: surveillance capabilities will expand, and entrepreneurs anticipate problems before they become acute.

This discovery process explains why privacy technology often leads user demand instead of following it. Entrepreneurs see what users do not yet recognize.

Creative Destruction

Schumpeter described innovation as creative destruction: new solutions displacing old ones, often destroying established businesses while creating new value.[11]

Privacy technology exhibits creative destruction. Encrypted messaging displaces SMS,[12] which is both less secure and more surveilled. Cryptocurrency displaces some payment systems built on surveillance and control.[13] Decentralized platforms compete with centralized ones whose revenue depends on data extraction.

This destruction is creative because it serves users better. The displaced solutions were inferior; their displacement is progress. But it is destructive for established interests that benefited from the old arrangements.

Entrepreneurial Judgment Under Uncertainty

Peter Klein emphasizes that entrepreneurship involves judgment under uncertainty:[14] decisions about resource allocation when outcomes cannot be known in advance.

Privacy entrepreneurs face substantial uncertainty. Adoption depends on factors they cannot control, and the legal environment may shift beneath them mid-development. Competitors may develop superior solutions while they build, and implementation itself may reveal problems no design anticipated.

Entrepreneurial judgment involves committing resources despite this uncertainty. Successful entrepreneurs are not those who eliminate uncertainty but those who judge well amid it.

8.4 Capital Heterogeneity and Privacy Specialization

Hayek on Capital Heterogeneity

Hayek showed that capital goods are heterogeneous: different tools serve different purposes and cannot be arbitrarily substituted.[15] Ludwig Lachmann deepened this analysis, showing that capital goods stand in relationships of complementarity and substitutability that only entrepreneurial judgment can assess.[16]

A hammer is not interchangeable with a saw, and a truck is not interchangeable with a ship; each capital good has specific uses and capabilities.

Investment decisions must therefore match capital to purpose. Misallocated capital cannot be reassigned, and coordination requires matching specific capitals to specific needs.

Privacy Capital Specialization

Privacy infrastructure exhibits heterogeneous capital characteristics. Cryptographic libraries cannot substitute for user-interface frameworks, and anonymous-network routing cannot substitute for identity management; each capital good has its function, and a well-developed library does not compensate for poor operational security.

Market Coordination of Heterogeneous Capital

Markets coordinate heterogeneous capital through price signals and entrepreneurial discovery:

Profit opportunities signal where capital is needed; high returns in a segment attract investment while low returns signal oversupply. Specialization enables expertise development as developers focus on specific components where they have comparative advantage. Exchange enables coordination between specialists, as those with cryptographic expertise trade with those having user interface expertise.

No central planner could coordinate this heterogeneous capital. The knowledge required, about specific capabilities, compatibility constraints, and market needs, is too dispersed. Only market processes can coordinate effectively.

8.5 Capital Formation and Privacy Development

The Capital Formation Process

Capital formation requires savings, investment, time, and maintenance.[17] Privacy capital formation follows this pattern:

In the savings phase, developers forgo current income to invest in infrastructure development, and users forgo convenient surveilled services to invest in learning privacy tools. In the investment phase, resources flow into protocol development, application creation, and network building; time and effort produce capital goods. The time dimension is substantial: privacy infrastructure takes years to develop and deploy, and returns in the form of privacy protection emerge gradually. Maintenance through ongoing development then preserves security properties; code review and updates maintain capital value.

Open Source as Capital Formation

Open source development is a distinctive form of capital formation:[18]

Open source involves collective investment, as many developers contribute without direct compensation to create shared capital. The resulting code is non-rivalrous capital: unlike physical capital, open-source code can be used by unlimited parties without depletion. Competitive improvement accelerates capital formation as multiple developers enhance the same code base. Transparency also increases capital quality because open code enables security verification.

This mode of capital formation enables privacy infrastructure development that proprietary approaches could not match. The capital is collectively created and freely available, and because the code is open it is continuously improved by anyone with the skill and motivation to contribute.

Capital Accumulation and Privacy Capability

Capital accumulates through successful investment:

Individual accumulation occurs as developers who build successful tools can use them for future projects; expertise compounds and networks of collaborators develop. System accumulation occurs as successful protocols enable new applications, infrastructure investment creates platforms for future development, and each layer enables the next. Knowledge accumulation occurs as solutions to

past problems inform future development; lessons learned become embedded in practices and code.

This accumulation explains why privacy technology has improved dramatically over decades. Each generation builds on previous achievements, using accumulated capital for further advancement.

Chapter Summary

Privacy infrastructure is capital in the Austrian sense, produced means of production that require present sacrifice for future capability. Roundabout methods of achieving privacy, which build cryptographic tools and protocols, outperform direct methods that rely on institutional promises: the second approach depends on the continuing goodwill of institutions the user cannot control, and the first does not. Higher-order goods (cryptographic foundations, protocols) enable lower-order goods (applications, services), and each generation of development compounds on the accumulated capital of the last.

Time preference theory explains why users split over the privacy-convenience tradeoff without either side being irrational. High time preference favors immediate convenience through surveillance-enabled services; low time preference favors investment in privacy infrastructure whose payoff accumulates over time. Markets coordinate these preferences through specialization, with low-time-preference developers building tools that high-time-preference users can later adopt. Entrepreneurial discovery drives the innovation itself: alert entrepreneurs notice unmet privacy needs and develop solutions that displace inferior surveilled systems. Capital heterogeneity means different tools serve different purposes, and no central planner could coordinate the dispersed knowledge their allocation requires.

The chapter provides the capital-theoretic foundation. Which specific tools to adopt, and how they work in practice, belong to Parts V and VI. The ethical case against state channels rests on Chapter 4's argumentation ethics and Chapter 5's analysis of coercion. The

institutional forms through which privacy capital is deployed at scale come in Chapters 24 through 26, and this chapter supplies the foundation they presuppose.[19]

Notes

[1]Eugen von Böhm-Bawerk, *Capital and Interest*, vol. 2: *Positive Theory of Capital*, trans. George D. Huncke and Hans F. Sennholz (South Holland, IL: Libertarian Press, 1959), 13–21.

[2]Israel M. Kirzner, *Competition and Entrepreneurship* (Chicago: University of Chicago Press, 1973), 35–42.

[3]Roger Dingledine, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium* (2004), `https://svn.torproject.org/svn/projects/design-paper/tor-design.pdf`; Tor Project, `https://www.torproject.org`. Tor is the canonical example of anonymous-network infrastructure as a capital good: years of academic research and ongoing development produce a routing layer that privacy applications depend on, but that no single application use exhausts.

[4]Bitcoin Core, `https://bitcoincore.org/`, with the original protocol description in Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008), `https://bitcoin.org/bitcoin.pdf`. For the systems-cluster treatment of Bitcoin in the cypherpunk lineage, see Chapter 2, note 8; for the security-model treatment, see Chapter 5, note 9.

[5]Carl Menger, *Principles of Economics*, trans. James Dingwall and Bert F. Hoselitz (Auburn, AL: Ludwig von Mises Institute, 2007 [1871]), 55–67. Menger's theory of goods orders, developed in the early chapters of *Principles*, provides the foundation for Austrian capital theory; the subjective-value treatment cited at Chapter 3, note 5 covers the later chapters of the same volume.

[6]The layered structure maps onto the protocol stack described in Steven Levy, *Crypto: How the Code Rebels Beat the Government — Saving Privacy in the Digital Age* (New York: Viking, 2001). Levy's narrative traces how academic public-key cryptography (highest-order) enabled PGP (mid-order), which enabled encrypted email applications (lowest-order), each layer depending on the one above and enabling the ones below.

[7]On cryptographic research timelines and the gap between theoretical foundations and deployed applications, see Whitfield Diffie and Martin E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* 22, no. 6 (1976): 644–654, `https://ieeexplore.ieee.org/document/1055638`. The 1976 paper established public-key cryptography as a theoretical possibility; RSA followed in 1977; mass deployment in TLS took until the mid-1990s, a gap of nearly two decades illustrating the higher-order/lower-order production structure described here.

[8]Böhm-Bawerk, *Capital and Interest*, vol. 2, chapters 1–4; see also "The Roundabout Process in the Interest Theory," *Quarterly Journal of Economics* 21, no. 4 (1907): 529–590.

[9]Hans-Hermann Hoppe, *Democracy: The God That Failed: The Economics and Politics of Monarchy, Democracy, and Natural Order* (New Brunswick, NJ: Transaction Publishers, 2001), chapters 1–2. Hoppe derives time preference theory's implications for capital accumulation and civilization from Mises's framework; his analysis explains why high-time-preference institutional environments systematically crowd out the long-horizon investment that privacy infrastructure requires.

[10]Kirzner, *Competition and Entrepreneurship*, 65–85. For development, see Kirzner, *Perception, Opportunity, and Profit: Studies in the Theory of Entrepreneurship* (Chicago: University of Chicago Press, 1979).

[11]Joseph A. Schumpeter, *Capitalism, Socialism and Democracy*, 3rd ed. (New York: Harper & Row, 1950), 82–85, where the "creative destruction" formulation is given its canonical statement.

[12]On SMS's security weaknesses, see the GSM Association's own acknowledgment of SS7 protocol vulnerabilities: Adrian Lison, "SS7 Vulnerabilities and the Threats They Pose to Mobile Network Security," *GSMA Intelligence* (2017); Karsten Nohl et al., "Roaming Off the Beaten Track: Cellular Traffic Interception," *31st Chaos Communication Congress* (2014). For the displacement by encrypted messaging, see Open Whisper Systems (now Signal Messenger), `https://signal.org`, and the Signal Protocol specification: Trevor Perrin and Moxie Marlinspike, "The Double Ratchet Algorithm" (2016), `https://signal.org/docs/specifications/doubleratchet/`.

[13]Friedrich A. Hayek, *Denationalisation of Money: The Argument Refined*, 2nd ed. (London: Institute of Economic Affairs, 1978). Hayek argued that government monopoly on money issuance is neither economically necessary nor beneficial, and that competing private currencies would serve consumers better, providing the theoretical foundation for understanding cryptocurrency as displacing surveillance-dependent payment systems on economic grounds.

[14]Peter G. Klein, *The Capitalist and the Entrepreneur: Essays on Organizations and Markets* (Auburn, AL: Ludwig von Mises Institute, 2010), 1–35.

[15]Friedrich A. Hayek, *The Pure Theory of Capital* (Chicago: University of Chicago Press, 1941), chapters 15–18; see also Hayek, *Individualism and Economic Order* (Chicago: University of Chicago Press, 1948), chapters 4–6.

[16]Ludwig M. Lachmann, *Capital and Its Structure* (London: Bell & Sons, 1956; reprinted Auburn, AL: Ludwig von Mises Institute, 1978), `https://mises.org/library/capital-and-its-structure`, chapters 2–3. Lachmann showed that capital goods are heterogeneous in kind and stand in complex relationships of complementarity and substitutability that vary with the production plan; no engineering specification can substitute for the entrepreneurial judgment that determines which combinations are viable.

[17]Peter Lewin, *Capital in Disequilibrium: The Role of Capital in a Changing World* (London: Routledge, 1999), 85–112. Lewin shows how capital formation involves knowledge accumulation.

[18]On open source as economic phenomenon, see Eric S. Raymond, *The Cathedral and the Bazaar*, rev. ed. (Sebastopol, CA: O'Reilly Media, 2001); Yochai Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom* (New Haven: Yale University Press, 2006).

[19]Further reading on Austrian capital theory. The classical Austrian treatment is Eugen von Böhm-Bawerk, *Capital and Interest*, 3 vols. (cited in note 1 above); Ludwig M. Lachmann, *Capital and Its Structure* (Sheed Andrews and McMeel, 1956; reprinted by Ludwig von Mises Institute, 1978), `https://mises.org/library/capital-and-its-structure`, develops the heterogeneity and complementarity of capital goods; Friedrich A. Hayek, *The Pure Theory of Capital* (cited in note 9 above), is the most rigorous extension. For the Kirznerian synthesis of capital, time, and entrepreneurial discovery, see Israel M. Kirzner, *An Essay on Capital* (Augustus M. Kelley, 1966); *Competition and Entrepreneurship* (cited in note 2 above) extends the framework, and Kirzner is also discussed at Chapter 7, notes 6 and 12. For modern Austrian treatments, Peter Lewin, *Capital in Disequilibrium* (cited in note 10 above); Peter Lewin and Nicolás Cachanosky, *Austrian Capital Theory: A Modern Survey of the Essentials* (Cambridge University Press, 2019), is the best contemporary entry point. On the specific question of time preference, Hans-Hermann Hoppe, *Democracy: The God That Failed* (cited at Chapter 5, note 12); Hoppe, *The Great Fiction* (Laissez Faire Books, 2012), develops the political-economy implications. On open-source production as a capital-formation mode, Yochai Benkler, *The Wealth of Networks* (cited in note 11 above); Steven Weber, *The Success of Open Source* (Harvard University Press, 2004), is the standard sociological treatment. For the economics of software and protocol development, Glyn Moody, *Rebel Code: Linux and the Open Source Revolution* (Perseus, 2001); Eric Raymond, *The Cathedral and the Bazaar* (cited in note 11 above), is the narrative history.

9

MONETARY THEORY AND SOUND MONEY

> "Money usually depreciates when it becomes too abundant."
>
> Nicolaus Copernicus, *Monete Cudende Ratio* (1517)[1]

Every exchange has two sides, and money is one of them. A theory of privacy in economic life has to start with money, not because money is the most interesting case, but because it is the unavoidable one. Whatever properties the monetary medium has, surveillance or privacy, traceability or opacity, permission-based or permissionless, those properties apply to every transaction conducted in it.

This chapter develops the Austrian theoretical framework for evaluating monetary systems. Money emerges through market process, not government decree: Menger showed why, and the argument has not been refuted. Sound money has specific properties derived from its three functions: medium of exchange, store of value, unit of account. Fiat money systematically undermines those properties while adding a fourth function its architects do not advertise: surveillance infrastructure. Digital money raises the question of whether those lost properties can be restored.

The chapter works through the theoretical foundations in sequence: the market origin of money, the regression theorem and its contested application to novel digital assets, the properties sound money requires (including transaction privacy, which standard

treatments omit), fiat money's systematic failures, the free banking alternative, the distinction between money proper and money substitutes, and what sound digital money would require. Chapters 18 through 19 apply this framework to Bitcoin specifically; the regression theorem debate and Bitcoin's privacy limitations are deferred to those chapters.[2]

9.1 The Market Origin of Money

Menger's Discovery

Carl Menger showed that money emerges through spontaneous market process.[3] No government invented money. No social contract established it. Money arose through the independent actions of individuals seeking to enable exchange.

The problem is barter's double coincidence of wants.[4] For direct exchange to occur, each party must want exactly what the other offers. A baker wanting shoes must find a shoemaker wanting bread. As economies develop and specialization increases, finding such coincidences becomes increasingly difficult.

The solution is indirect exchange. Individuals begin accepting goods they do not directly want but know they can trade further. A baker might accept cloth not to wear but because the shoemaker will accept cloth. The baker trades bread for cloth, then cloth for shoes, accomplishing through two exchanges what direct barter could not achieve.

Emergence of Money

Certain goods prove especially useful for indirect exchange. Menger identified the property of "salability": how readily a good can be exchanged for other goods. Highly saleable goods become preferred media of exchange.

Salability depends on divisibility (can the good be divided for small transactions?), durability (does the good maintain value over time?), portability (can the good be transported efficiently?), and recognizability (can the good be readily identified and verified?).

Goods with superior salability, historically precious metals, become generally accepted media of exchange. Money emerges: not by decree but by market selection.

Stages of Monetary Development

Menger identified a progression from direct barter (where exchange is limited by double coincidence of wants) through indirect exchange (where superior goods are accepted for further trade) to general acceptance (where the most saleable good becomes universally accepted) and finally to unit of account status (where money becomes the standard for pricing and calculation). This progression occurs through individual choices, not collective decision. Each person, seeking easier exchange, gravitates toward goods others will accept. The result is spontaneous convergence on money.

9.2 The Regression Theorem

The Problem of Monetary Value

How does money acquire value? Ordinary goods are valued for their direct use, but money is valued for its exchange power, which itself depends on money's value. The reasoning seems circular.

Mises resolved this through the regression theorem:[5] Money's current value depends on its expected future purchasing power, which depends on yesterday's purchasing power. The chain regresses to the point when the money commodity was valued for non-monetary uses.

Gold was valued for jewelry and ornament before it was money. Silver had industrial and decorative uses. Cattle served as food and labor before serving as money. Each successful money began as a commodity valued for direct use, then acquired additional monetary demand through its salability.

Application to Novel Moneys

The regression theorem raises questions about new moneys that never had non-monetary use. If money must trace back to prior com-

modity value, how can a novel money emerge? This question has divided Austrian economists, and the debate remains unresolved.[6]

Some scholars argue that novel digital moneys cannot satisfy the theorem. On this view, money must emerge from a commodity with prior use-value; digital assets lacking physical commodity backing cannot become true money but only function as media of exchange built on top of existing fiat money.

The opposing view holds that the theorem's requirements are satisfied by subjective valuation of any kind. The subjective theory of value, foundational to Austrian methodology, holds that value exists only in the minds of valuing individuals. The theorem establishes that money's current value traces back to prior valuations, but those prior valuations themselves derive from individual subjective assessments, not from objective commodity properties. If first valuers had any reason for valuing, and that value was transmitted through market exchange, the regression chain can begin. Praxeology provides no basis for declaring some subjective valuations legitimate and others illegitimate.

A further interpretive point separates the theorem's explanatory power from metaphysical necessity. The theorem explains how money typically emerges through market process, showing that monetary value need not be decreed by authority. It does not necessarily restrict which goods can become money if market participants choose to value them.

This book adopts the subjective value interpretation. The question remains contested within Austrian economics, and thoughtful scholars disagree. The theorem's core insight, that money emerges through market process and not by decree, is not in dispute. The open question is whether digital assets with no prior physical commodity use can satisfy the theorem's requirements, or whether they constitute a novel phenomenon that expands monetary theory. Chapter 19 applies this framework to Bitcoin specifically.

9.3 Sound Money Properties

Sound money has properties enabling it to serve economic coordination. These properties derive from money's functions, not from arbitrary preference.[7]

The Three Classical Functions

Money serves three functions, and each function imposes requirements on the medium that performs it.

As a **medium of exchange**, money must be acceptable in trade. Trading partners need to verify authenticity without specialized equipment, which requires recognizability and makes counterfeiting costly. The medium must subdivide for transactions of any size, since indivisible money limits exchange. And it must be transportable relative to its value; a high value-to-weight ratio enables larger transactions and reaches farther.

As a **store of value**, money must preserve purchasing power over time. Physical money must resist degradation and digital money must resist data loss, which together amount to durability. Supply must be predictable, because arbitrary inflation destroys the store-of-value function regardless of the medium's other merits. And the holder must be able to resist confiscation, because money that can be seized at will offers poor storage no matter what its monetary properties are in the abstract.

As a **unit of account**, money must enable economic calculation. Value must be reasonably stable for prices to convey information; wild fluctuations corrupt the price signals that coordinate production and exchange. Units must be interchangeable, because money whose value varies with its history complicates calculation at every step and defeats it.

Privacy as a Sound-Money Property

Transaction privacy is a sound-money property often overlooked in standard treatments but implied by the others. As Chapter 7 established, exchange functions best when parties control disclosure; money that exposes all transactions distorts voluntary coordination.

Money that enables third-party monitoring also enables third-party interference, and money whose units can be distinguished and discriminated against loses the fungibility that calculation requires. Privacy is therefore not a separate desideratum bolted onto the medium-of-exchange function; it is a condition under which the other three functions continue to work.

Physical cash exhibits this property natively: transactions leave no automatic record. The absence of records is not a bug but a feature enabling voluntary exchange without surveillance. Digital money that forfeits this property forfeits more than privacy. It forfeits some of the monetary capacity the privacy protected.

9.4 Fiat Money and Its Problems

What Fiat Money Is

Fiat money is money by government decree. It has no commodity backing. Its value derives from legal tender laws requiring its acceptance and from tax obligations payable only in fiat currency.

Modern fiat currencies (dollar, euro, yen) exemplify this: paper and digital entries with value because governments say so and enforce accordingly.

The Inflation Problem

Fiat money enables unlimited supply expansion. Central banks create money at will, subject only to political constraints. This creates systematic problems.[8]

New money enters the economy through specific channels such as bank lending and government spending. First recipients spend at old prices; later recipients face inflated prices. This mechanism transfers wealth from later to earlier recipients. When prices rise unpredictably, businesses cannot distinguish real demand changes from monetary distortion, and resources flow to inflation-favored sectors instead of consumer-preferred uses, corrupting economic calculation. Inflation punishes savers whose purchasing power declines while rewarding debtors whose obligations shrink in real

terms, thereby discouraging capital accumulation. Artificial credit expansion creates unsustainable booms followed by necessary busts; malinvestment during expansion must be liquidated during contraction, generating the business cycle.

The Surveillance Problem

Fiat money in its modern form is surveilled money. Digital fiat transactions are recorded, reported, analyzable, and controllable: banks maintain complete histories as a matter of ordinary operation, regulations require disclosure to government agencies at falling thresholds, pattern analysis extracts personal information from the resulting flows, and accounts can be frozen or transactions blocked at the direction of authorities the holder does not choose.

This surveillance capability has expanded dramatically through Know Your Customer (KYC) requirements,[9] Anti-Money Laundering (AML) monitoring, automatic information sharing between jurisdictions,[10] and lower transaction reporting thresholds that capture ever more transactions.

The result is simple: fiat money use generates full surveillance records. Privacy in economic life requires alternatives to surveilled fiat systems.

The Control Problem

Fiat money enables economic control at several scales. Authorities can freeze accounts and prevent individuals from accessing their own money; they can block specific payments; they can deny financial access entirely through deplatforming; and through sanctions they can exclude entire nations from payment networks. Each capability rests on the same fact: the money is a claim on an intermediary the state can reach.

This control can be used against criminals but also against dissidents, journalists, activists, and anyone disfavored by those with control. Fiat money is permission-based money: you may transact if authorities allow.

9.5 Free Banking vs. Central Banking

The Free Banking Alternative

Free banking is competitive money and banking without central bank monopoly.[11] Multiple banks issue their own notes, competing for customers based on reliability and service.

Historical examples (Scotland 1716–1845, Canada before 1935) suggest free banking systems were more stable than central banking, with fewer crises and better customer service. This assessment, associated with Lawrence White and George Selgin, is contested within Austrian thought; Murray Rothbard and others favored 100% reserve banking over fractional reserve free banking.[12] Market discipline prevented excessive risk-taking; banks that overissued lost customers to more conservative competitors.

Free banking preserves privacy through competition. Banks serve customers; customers who value privacy choose privacy-respecting banks. No single authority can impose surveillance across all financial activity.

Central Banking Problems

Central banking replaces market competition with monopoly, and several problems follow from the replacement. A single point of failure emerges, because one institution's errors affect the entire economy at once. Political capture follows, because monetary policy then serves political ends instead of economic ones. Moral hazard develops when banks take excessive risks knowing central banks will bail them out. And a single regulatory framework enables standardized surveillance and full monitoring.

The transition from competitive to central banking, completed in most countries by mid-twentieth century, replaced market-based money with politically controlled money, with corresponding privacy losses.

9.6 Money Proper vs. Money Substitutes

Austrian monetary theory distinguishes between money proper and money substitutes.[13]

Money proper is base money: the final means of payment requiring no further redemption. Gold coins in hand are money proper. You possess the value directly. No issuer exists to trust, no counterparty who must perform, no claim to be honored.

Money substitutes are claims against money proper. A bank note promising gold on demand is a money substitute. A bank account balance is a money substitute. The holder possesses not the money itself but a promise from an issuer.

Money substitutes carry counterparty risk. The holder must trust that the issuer exists, holds the reserves promised, will honor redemption requests, and cannot be prevented from honoring them. When issuers fail, are shut down, or refuse redemption, money substitutes become worthless regardless of whether the underlying money proper still exists.

This distinction matters for digital money. Every digital currency before Bitcoin was a money substitute: a claim against an issuer maintaining account balances. E-gold accounts were claims on gold held by a company, Liberty Reserve balances were claims against that company's bookkeeping, and DigiCash tokens were claims validated by a central server. When these issuers were shut down, users lost everything.[14] The failure was inherent: money substitutes require trusted issuers, and trusted issuers are vulnerable.

The challenge for sound digital money: can a digital asset be money proper, not a money substitute? Can it be the thing itself, not a claim?

The Current Monetary Architecture

The money proper/money substitutes distinction illuminates today's fiat monetary structure.

Base money (money proper) exists in two forms. Physical cash is base money that the general public can hold directly. Central bank reserves are also base money, but they are accessible only to

commercial banks and governments. Ordinary citizens cannot open accounts at the Federal Reserve[15] or European Central Bank.[16]

What citizens hold in bank accounts is not base money but money substitutes: claims against commercial banks. Your bank balance is an IOU from the bank, not base money itself. This is why bank failures destroy depositors' balances even though the underlying base money still exists.

This architecture creates a buffer between state and citizen. The central bank issues base money to commercial banks; commercial banks issue money substitutes to citizens. Citizens interact with the monetary authority only indirectly, through private institutions. Chapter 10 examines how Central Bank Digital Currencies would eliminate this buffer by giving citizens direct base money balances at the central bank, with direct consequences for surveillance and control.

9.7 Digital Money Requirements

What would sound digital money require? Austrian monetary theory suggests several criteria.[17]

Sound digital money requires decentralized verification: solving the double-spending problem without trusted third parties, with no single point that can be captured, corrupted, or coerced, and with verification distributed among participants instead of concentrated in one authority.

Digital scarcity through rivalrousness is essential for property rights to apply, and it is where digital money diverges from ordinary digital information. Units must be rivalrous, which means one person's possession excludes another's. Physical goods are naturally rivalrous, while digital information is not. Solving double-spending creates rivalrousness: if the same token cannot be spent twice, possession becomes exclusive, and only then can property rights emerge. Without rivalrousness, digital tokens remain information, freely replicable and incapable of serving as property.

Distinct from scarcity itself is the requirement for transparent and immutable supply. Sound digital money requires that total

supply and issuance rules be transparent, verifiable, and not subject to unilateral change by any third party. This differs from fiat money, where supply decisions are opaque and policy changes sit at central bank discretion. Transparency enables rational economic calculation in ways opaque monetary policy cannot.

User-defined rules invert the traditional model where money systems impose rules from above and users comply or leave. In sound digital money, each participant defines, verifies, and enforces the rules they accept. Running a full node means validating every transaction against self-chosen rules. Consensus emerges from participants independently converging on compatible rules, not from authority imposing uniformity. The rules are what the network enforces, not what some authority declares. Changes require convincing users to adopt new software; the default is the status quo.

Sound digital money must also be permissionless, privacy-preserving, censorship-resistant, and voluntarily adopted. Permissionless access means no gatekeeper controls use. Transaction privacy means verification is possible without automatic surveillance, with users controlling what is revealed. Censorship resistance means transactions execute whenever the protocol's rules are met; no third party can permit or deny. Voluntary adoption means the system emerges through market process and validates its value through exchange.

9.8 Bridge to Bitcoin

Chapters 18 through 19 examine Bitcoin as an implementation of these requirements. Here we note the basic contours of that analysis.

Bitcoin attempts to satisfy monetary theory requirements through decentralized verification via proof-of-work consensus,[18] digital scarcity through solving double-spending, transparent and immutable supply through protocol rules enforced by every node, and user sovereignty through each participant independently validating the rules they accept. Bitcoin does have privacy limitations: basic Bitcoin transactions are pseudonymous but not

anonymous, chain analysis can link transactions to identities,[19] and privacy requires additional tools discussed in Chapter 20.

Chapter 19 also develops the regression theorem analysis for Bitcoin specifically, examining how the subjective value framework applies to its emergence as money.

Chapter Summary

Money emerges through market process, not government decree. Menger showed that individuals seeking to overcome barter's double coincidence of wants naturally converge on goods with superior salability, and the spontaneous emergence explains money's origin without central planning. Sound money has properties derived from its three functions, medium of exchange, store of value, unit of account, and transaction privacy belongs with the others: money that exposes all transactions distorts the voluntary coordination the other properties exist to support.

Fiat money creates systematic problems the Austrian tradition has documented for a century. Unlimited supply expansion transfers wealth from later to earlier recipients, corrupts the calculation on which production depends, destroys savings, and generates the business cycle. Modern fiat is also surveillance money, generating transaction records that enable monitoring and control. The distinction between money proper and money substitutes illuminates the current architecture: physical cash is the only base money citizens can hold directly, while account balances are claims on commercial banks that can be frozen, seized, or denied under legal pressure the bank cannot refuse. Chapter 10 examines what happens when the state tries to remove that buffer through central bank digital currencies.

Digital money can restore the properties fiat has lost. The requirements, decentralized verification, digital scarcity through rivalrousness, transparent and immutable supply, user-defined rules, permissionless access, transaction privacy, censorship resistance, are what Chapters 18 through 20 assess for Bitcoin specifically, including its real privacy limitations. The regression theorem's

application to novel digital money remains contested within Austrian economics; this book adopts the subjective-value interpretation while leaving the underlying question open. And the chapter critiques fiat's systematic problems while granting its medium-of-exchange function.[20]

Notes

[1]Nicolaus Copernicus, *Monete Cudende Ratio* (*On the Minting of Money*, 1517; sometimes titled *Monetae Cudendae Ratio*). English translation in Nicholas Copernicus, *Minor Works*, trans. Edward Rosen and Erna Hilfstein (Baltimore: Johns Hopkins University Press, 1992; Warsaw: Polish Scientific Publishers, 1985). Written decades before Sir Thomas Gresham formulated "bad money drives out good," Copernicus identified both the quantity-theory principle behind currency debasement and what would later be called Gresham's Law. For the canonical modern Austrian treatment the chapter draws on, see Mises, *Human Action* (cited at Chapter 3, note 1), 398 and part IV.

[2]Accessible book-length treatments covering the arc from commodity money to fiat to digital money: Saifedean Ammous, *The Bitcoin Standard* (Hoboken, NJ: John Wiley & Sons, 2018); Nik Bhatia, *Layered Money* (Self-published, 2021); Lyn Alden, *Broken Money: Why Our Financial System Is Failing Us and How We Can Make It Better* (Timestamp Press, 2023). Each approaches the same arc from a different angle: Ammous from Austrian monetary theory, Bhatia from the layered structure of monetary systems, Alden from monetary history and macro.

[3]Carl Menger, "On the Origin of Money," *Economic Journal* 2, no. 6 (1892): 239–255. See also Menger, *Principles of Economics* (cited at Chapter 3, note 5), chapter 8.

[4]W. Stanley Jevons, *Money and the Mechanism of Exchange* (London: Kegan Paul, 1875), chapter 1. Jevons coined the phrase "double coincidence of wants" and gave the first systematic account of why barter fails at scale; Menger's independent discovery of money's market origin (cited in note 3 above) provides the complementary supply-side analysis.

[5]Ludwig von Mises, *The Theory of Money and Credit*, trans. H. E. Batson (Indianapolis: Liberty Fund, 1981 [1912]), 97–123. See also Mises, *Human Action* (cited at Chapter 3, note 1), 408–416.

[6]For skeptical views on Bitcoin and the regression theorem, see Frank Shostak, "The Bitcoin Money Myth," *Mises Daily* (April 17, 2013), `https://mises.org/mises-daily/bitcoin-money-myth`. For defenses of Bitcoin's compatibility with the theorem, see Peter Šurda, "Economics of Bitcoin: Is Bitcoin an Alternative to Fiat Currencies and Gold?" (master's thesis, Vienna University of Economics and Business, 2012); Konrad Graf, "On the Origins of Bitcoin: Stages of Monetary Evolution" (2013), `https://konradsgraf.com`; Eric Voskuil, "Regression Fallacy," in *Cryptoeconomics* (cited at Chapter 5, note 1),

https://github.com/libbitcoin/libbitcoin-system/wiki/Regression-Fallacy. For a recent synthesis, see Laura Davidson and Walter Block, "Bitcoin, the Regression Theorem, and the Emergence of a New Medium of Exchange," *Quarterly Journal of Austrian Economics* 18, no. 3 (2015): 311–338.

[7]On sound money properties, see Mises, *The Theory of Money and Credit* (cited in note 4 above), Part I; Murray N. Rothbard, *What Has Government Done to Our Money?*, 5th ed. (Auburn, AL: Ludwig von Mises Institute, 2010), chapters 1–3.

[8]On fiat money problems, see Rothbard, *What Has Government Done to Our Money?* (cited in note 6 above); Jörg Guido Hülsmann, *The Ethics of Money Production* (Auburn, AL: Ludwig von Mises Institute, 2008); Hans-Hermann Hoppe, "How Is Fiat Money Possible? or, The Devolution of Money and Credit," *Review of Austrian Economics* 7, no. 2 (1994): 49–74.

[9]Bank Secrecy Act, Pub. L. 91-508, 84 Stat. 1114 (1970), codified at 31 U.S.C. §§ 5311–5336. The BSA established the foundational KYC and AML reporting framework in the United States, requiring financial institutions to maintain transaction records and file Currency Transaction Reports (CTRs) for cash transactions above $10,000 and Suspicious Activity Reports (SARs) for unusual patterns. For current FinCEN implementation guidance, see U.S. Department of the Treasury, Financial Crimes Enforcement Network, https://www.fincen.gov.

[10]Financial Action Task Force (FATF), *International Standards on Combating Money Laundering and the Financing of Terrorism and Proliferation* (Paris: FATF, 2012, updated 2023), https://www.fatf-gafi.org/en/topics/fatf-recommendations.html; OECD, *Standard for Automatic Exchange of Financial Account Information in Tax Matters*, 2nd ed. (Paris: OECD Publishing, 2017), https://doi.org/10.1787/9789264267992-en. The FATF Recommendations set the global KYC/AML standard adopted by over 200 jurisdictions; the OECD Common Reporting Standard (CRS) operationalizes automatic cross-border financial account data exchange, giving states full visibility into citizens' foreign holdings.

[11]On free banking, see Lawrence H. White, *Free Banking in Britain: Theory, Experience and Debate, 1800–1845*, 2nd ed. (London: Institute of Economic Affairs, 1995); George A. Selgin, *The Theory of Free Banking* (Lanham, MD: Rowman & Littlefield, 1988).

[12]For the Rothbard critique of fractional reserve free banking, see Murray N. Rothbard, *The Mystery of Banking*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2008), and "The Case for a 100 Percent Gold Dollar," in *In Search of a Monetary Constitution*, ed. Leland Yeager (Cambridge, MA: Harvard University Press, 1962). The debate between 100% reserve advocates (Rothbard, Jesús Huerta de Soto) and fractional reserve free bankers (White, Selgin) remains unresolved within the Austrian tradition.

[13]On money proper vs. money substitutes (also called fiduciary media), see Mises, *The Theory of Money and Credit* (cited in note 4 above), Part II, chapters 1–3; also Mises, *Human Action* (cited at Chapter 3, note 1), 432–444.

[14]E-gold operated from 1996 until indictment in 2007 and shutdown in 2009 (*United States v. E-Gold, Ltd.*, D.D.C. 2008). Liberty Reserve operated from 2006 until U.S. seizure in May 2013 (*United States v. Liberty Reserve S.A.*, S.D.N.Y. 2013). DigiCash, founded by David Chaum in 1989, filed for bankruptcy in 1998. On Chaum's earlier untraceable-payments work, see Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms" (cited at Chapter 2, note 6).

[15]Federal Reserve Act, Pub. L. 63-43, 38 Stat. 251 (1913). The Federal Reserve Act established the U.S. central banking system, creating the Federal Reserve Board and twelve regional Federal Reserve Banks. Reserve accounts at the Fed are accessible only to member commercial banks and certain other depository institutions, not to the general public. For the Federal Reserve's current account access policy, see Board of Governors of the Federal Reserve System, `https://www.federalreserve.gov`.

[16]Treaty on the Functioning of the European Union, art. 282 (2007); Protocol No. 4 on the Statute of the European System of Central Banks and of the European Central Bank. The ECB was established by the Treaty of Maastricht (1992) and began operations in 1998; like the Federal Reserve, it maintains reserve accounts for eurozone credit institutions but not for natural persons or non-financial entities. European Central Bank, `https://www.ecb.europa.eu`.

[17]For application of Austrian monetary theory to digital money, see Ammous, *The Bitcoin Standard* (cited in note 2 above); Eric Voskuil, *Cryptoeconomics: Fundamental Principles of Bitcoin* (cited at Chapter 5, note 1).

[18]Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008), `https://bitcoin.org/bitcoin.pdf`. Section 4 introduces proof-of-work as the mechanism for achieving distributed consensus without a trusted third party; the cumulative-work chain replaces the need for a central authority to order transactions and thereby solves double-spending. This is the foundational technical document for all subsequent analysis of Bitcoin as a monetary system.

[19]Sarah Meiklejohn et al., "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names," *Proceedings of the 2013 Internet Measurement Conference* (ACM, 2013), `https://dl.acm.org/doi/10.1145/2504730.2504747`. Meiklejohn et al. showed that Bitcoin's public blockchain allows heuristic clustering of addresses by common-input-ownership, enabling de-anonymization of a significant fraction of transactions through graph analysis, the foundational academic result establishing that pseudonymity differs materially from anonymity.

[20]Further reading on monetary theory, for the reader coming to this chapter from outside the Austrian tradition. The indispensable primary source is Mises, *The Theory of Money and Credit* (cited in note 4 above), the foundational Austrian treatment that first applied Menger's marginalism to money. Mises's later work *Human Action* (cited at Chapter 3, note 1), parts III–IV, integrates monetary theory with the full praxeological framework. Rothbard, *What Has Government Done to Our Money?* (cited in note 6 above), is the shortest introduction; *The Mystery of Banking* (cited in note 9 above) is the longer practitioner treatment. Jesús Huerta de Soto, *Money, Bank Credit, and Economic Cycles*, 3rd ed. (Auburn, AL: Ludwig

von Mises Institute, 2012), https://mises.org/library/money-bank-credit-and-economic-cycles, is the broadest contemporary Austrian treatment of money and banking. For the free-banking side, Lawrence H. White, *The Theory of Monetary Institutions* (Oxford: Blackwell, 1999), and Selgin, *The Theory of Free Banking* (cited in note 8 above) are the reference works. For narrative history, Murray N. Rothbard, *A History of Money and Banking in the United States* (Auburn, AL: Ludwig von Mises Institute, 2002), and Vera C. Smith, *The Rationale of Central Banking and the Free Banking Alternative* (Indianapolis: Liberty Fund, 1990 [1936]), cover the institutional history. For the monetary-regression side, see Menger, "On the Origin of Money" (cited in note 3 above), and its application to Bitcoin in Šurda, "Economics of Bitcoin" (cited in note 5 above), and Graf, "On the Origins of Bitcoin" (cited in note 5 above). For the opposing school worth serious engagement: Milton Friedman and Anna J. Schwartz, *A Monetary History of the United States, 1867–1960* (Princeton, NJ: Princeton University Press, 1963), is the foundational monetarist empirical work; Milton Friedman, *A Program for Monetary Stability* (New York: Fordham University Press, 1960), is the shortest policy statement; John Maynard Keynes, *A Tract on Monetary Reform* (London: Macmillan, 1923), provides the opposing tradition's monetary analysis; for contemporary heterodox monetary thought, Perry Mehrling, *The New Lombard Street: How the Fed Became the Dealer of Last Resort* (Princeton, NJ: Princeton University Press, 2011), explains the modern institutional structure that Austrian theory critiques.

PART IV

THE ADVERSARY

10

FINANCIAL SURVEILLANCE AND STATE CONTROL

> "The State is an organization of the political means."
>
> Franz Oppenheimer, *The State* (1907)[1]

Murray Rothbard's intervention typology distinguishes autistic and binary intervention from triangular intervention, in which the state imposes on third-party exchanges.[2] The framework illuminates how states surveil and control through finance, culminating in Central Bank Digital Currencies that combine all three intervention types.

10.1 Intervention Theory: Autistic, Binary, Triangular

Rothbard's Framework

Rothbard developed a systematic typology of government intervention in *Power and Market*.[2] All intervention involves the use or threat of violence to alter behavior that would otherwise occur in unhampered markets.

Autistic intervention involves commands directed at individuals without any exchange taking place. (Rothbard uses "autistic" in its technical economic sense, derived from Mises, meaning self-contained or non-exchange action, not in any psychological sense.) The state orders: do this, do not do that. No transaction occurs;

the individual is commanded. Examples include conscription and compulsory schooling, as well as prohibitions on consumption.

Binary intervention involves the state as a party to an exchange, typically involuntary from the other party's perspective. Taxation is the classic case: the state takes wealth from individuals, giving nothing in return that the individual would have voluntarily purchased. Asset seizure and eminent domain, along with confiscation, are binary interventions.

Triangular intervention involves the state imposing on exchanges between third parties. The state is not a party to the transaction but dictates its terms. Price controls and licensing requirements, along with mandatory contract terms, are triangular interventions. The state forces or forbids exchanges it does not participate in.

Application to Financial Privacy

This framework illuminates financial surveillance. Autistic intervention encompasses commands regarding privacy behavior, such as encryption bans or prohibitions on anonymous transactions. Binary intervention involves state extraction of information or assets through subpoenas and seizures, along with compelled disclosure. Triangular intervention imposes mandates on private transactions, such as KYC requirements forcing banks to collect customer information before providing services.

Most financial surveillance operates through triangular intervention. The state does not directly surveil citizens (which would require resources and face constitutional constraints). Instead, it forces private institutions to surveil on its behalf, bearing the costs while providing the data.

10.2 Autistic Intervention and Privacy

Direct Criminalization

Autistic intervention in privacy takes the form of direct prohibitions on tools and behaviors. U.S. export controls until 1996 classified cryptographic software as "munitions," treating mathematical

algorithms as weapons that required government permission to share, and some nations still restrict encryption strength or require key escrow.[3] Proposals to ban privacy coins, coinjoining[4] services, or anonymous communication tools fit the same category: the individual is commanded, and no exchange with the state occurs.

Economic Analysis

Autistic intervention in privacy faces inherent limitations. The state cannot easily detect privacy tool usage without the surveillance it seeks to impose; encryption looks like random data, and its presence is deniable. Direct prohibition requires identifying violators, which requires the surveillance capability that privacy tools defeat; the intervention is self-undermining. Digital tools cross borders, so prohibition in one jurisdiction shifts activity instead of eliminating it.

These limitations explain why states prefer triangular intervention for financial surveillance. Direct prohibition is resource-intensive and often ineffective against technically sophisticated users.

10.3 Binary Intervention and Privacy

State as Extracting Party

Binary intervention places the state as a party to involuntary exchange, extracting from individuals instead of imposing on third-party transactions.

Civil asset forfeiture enables seizure without criminal conviction. Property is accused of connection to crime; the owner must prove innocence to recover assets. This binary intervention extracts wealth while creating chilling effects on financial privacy.[5]

Court orders forcing individuals to reveal passwords or encryption keys, as well as account information, are another form of binary intervention. The state extracts information directly, using contempt sanctions to coerce compliance.

Government demands for records held by individuals (as opposed to third-party reporting requirements) are also binary interventions. The individual must surrender information to the state.

While taxation itself is binary (the state takes, the individual receives nothing), compliance requirements impose additional binary burdens. Individuals must disclose financial information to demonstrate compliance, revealing private activity regardless of tax liability.

Resistance and Limits

Binary intervention faces the target directly. Unlike triangular intervention operating through intermediaries, binary intervention requires the state to identify, locate, and coerce specific individuals.

In some jurisdictions, direct government extraction faces procedural requirements (warrants, probable cause) that triangular intervention circumvents. The Fourth Amendment constrains what the state can demand directly but not what third parties collect and report.[6]

Forced disclosure of encrypted data faces the practical problem that compliance cannot be verified. If the defendant claims to have forgotten a password, proving otherwise requires showing knowledge that, if the state possessed it, would make compulsion unnecessary.

Properly implemented encryption resists binary intervention entirely. The state can demand keys, but if keys were never stored or are forgotten, the demand cannot be satisfied. The Axiom of Resistance operates in practice here.[7]

10.4 Triangular Intervention and Privacy

The Preferred Mechanism

Triangular intervention is the primary mechanism of financial surveillance. Instead of directly commanding individuals or extracting from them, the state imposes on private transactions, forcing parties to modify their exchanges.

The Bank Secrecy Act of 1970 exemplifies this mechanism. Despite its name, the BSA destroyed bank secrecy by requiring financial institutions to maintain records and report transactions. Banks must file Currency Transaction Reports (CTRs) for cash transactions over $10,000 and Suspicious Activity Reports (SARs) when they know, suspect, or have reason to suspect that a transaction meets the applicable suspicious-activity standard.[8] This is triangular intervention: the state imposes on bank-customer relationships it is not party to. The bank must surveil customers; customers must submit to surveillance to access banking.

Know Your Customer mandates force financial institutions to collect and verify customer identity before establishing relationships. The customer-bank exchange is conditioned on information disclosure to satisfy government requirements. Anonymous purchase threshold laws follow the same structure: Alice and Bob wish to transact privately above a specified amount, and Charlie (the state) forbids it. The state is not a party to the exchange; it dictates terms to parties who would otherwise have agreed differently. The threshold triggers a disclosure obligation imposed on both buyer and seller, not a command directed at either in isolation. KYC and Customer Identification Program rules require specified identifying information for customers and, for legal-entity customers, beneficial-ownership information; depending on the customer's risk profile, institutions may also request proof of address or source-of-funds documentation. This information serves surveillance, not the bank's commercial interest. Banks would prefer simpler, less costly customer relationships; KYC is imposed externally.

Third-party reporting requirements compel financial institutions to report to government agencies: CTRs, SARs, FATCA filings for foreign accounts, Form 1099s for income. The individual's financial activity is reported without their consent, often without their knowledge.

Monopoly Grants as Triangular Intervention

Monopoly, in the praxeological sense, is a grant of exclusive privilege by the state, not a dominant market position achieved through

superior service. Rothbard distinguished between market dominance (where a single seller emerges through voluntary exchange) and monopoly privilege (where the state prohibits competition). Only the latter is monopoly proper.

Financial surveillance relies on monopoly grants. Banking licenses restrict who may accept deposits, creating a cartel that must obey surveillance mandates to maintain its privilege. Payment network regulations restrict who may transmit money, ensuring all major pathways are surveilled. Legal tender laws require acceptance of state currency, channeling transactions through monitored systems. Each monopoly grant creates a point of control: the state offers exclusive privilege, and the price is surveillance compliance.

The most fundamental monopoly grant is the exclusive right to issue the currency itself. The Federal Reserve holds the sole legal authority to create U.S. dollar-denominated base money.[9] The surveillance dimension is secondary to the fiscal one: any good produced under competitive conditions will see its price driven toward its marginal cost of production. Paper currency costs fractions of a cent per note to print. If competitors could legally issue dollar-denominated currency, competition would drive the exchange value of that currency toward its production cost, eliminating the gap between face value and cost (the gap the state captures as seigniorage). The currency monopoly is therefore the precondition for money printing as a revenue source. Open competition in currency issuance would extinguish seigniorage as surely as open competition in any other industry extinguishes above-cost returns. Legal tender laws and counterfeiting prohibitions together preserve this gap, ensuring that base-money creation remains a state fiscal instrument, insulated from competitive pressure.

This explains why privacy-preserving alternatives face legal attack. Bitcoin threatens the currency monopoly,[10] unlicensed money transmission threatens the payment monopoly, and privacy coins threaten the surveillance infrastructure built on these monopolies. The legal response is predictable: extend monopoly protections to exclude competition.

Why States Prefer Triangular Intervention

Triangular intervention offers advantages over autistic and binary approaches.

Banks bear surveillance costs; compliance infrastructure (personnel, technology, reporting systems) is privately funded but serves government objectives. Major banks spend billions annually on compliance, and these costs are ultimately borne by customers and shareholders.

Information collected by private parties under commercial authority faces fewer legal protections than information the government collects directly. Third-party doctrine holds that information shared with private parties loses constitutional protection. Banks serve millions of customers, so forcing banks to surveil achieves broad coverage that direct government surveillance could not match.

Rothbard observed that interventions create problems requiring further intervention, and Mises analyzed the underlying mechanism.[11] Interventionism is unstable because each intervention produces consequences it did not intend, and those consequences invite further intervention to address. At each step the state confronts a choice between withdrawing the original intervention and extending it; the political economy of administration produces extension, because the apparatus built for the first step is already paid for and staffed. Financial surveillance has followed this pattern across five decades. The Bank Secrecy Act of 1970 imposed reporting requirements that revealed gaps. The Money Laundering Control Act of 1986[12] closed some gaps by making the underlying conduct a predicate offense and adding Suspicious Activity Reports to the Currency Transaction Report regime. The USA PATRIOT Act of 2001[13] extended the regime to non-bank money-service businesses whose growth had been one of the revealed gaps. The 2016 Customer Due Diligence rule[14] required banks to identify beneficial owners behind legal-entity customers, closing the gap that shell companies had opened. The FATF Travel Rule extension to virtual-asset service providers in 2019 and its global rollout through 2024 closed the gap that non-custodial crypto transfers had opened. Each step was a

logical response to a gap the previous step exposed, and each step enlarged the surveillance apparatus that would expose the next gap. No equilibrium holds at any intermediate level: the retreat direction has no political constituency inside the apparatus, and the extension direction has a predictable one on every budget cycle.

Economic Effects

Triangular intervention distorts the transactions it targets.

Resources devoted to surveillance (personnel, technology, legal, reporting) are unavailable for productive activity. This is a deadweight loss: surveillance does not create value for the parties to the transaction. Banks, facing regulatory risk from "suspicious" customers, deny services to entire categories: cryptocurrency businesses, cannabis companies, politically disfavored groups, even individual account holders flagged by opaque algorithms. The intervention cascades into complete exclusion from financial services.

Knowledge of surveillance alters behavior. Market participants avoid legitimate transactions that might trigger reports, and economic coordination suffers as parties steer away from flagged activities. Financial privacy disappears as full transaction records accumulate in government databases. The "financial panopticon" makes economic activity legible to state observation.

10.5 Central Bank Digital Currencies as Total Intervention

From Triangular to Direct Intervention

Chapter 9 established that today's monetary architecture interposes commercial banks between citizens and base money. Citizens hold money substitutes (claims on banks), not base money itself. This buffer forces the state to work through intermediaries: banks collect data under regulatory compulsion, and the state accesses it through legal process. Financial surveillance currently operates primarily through triangular intervention.

In the strongest retail CBDC designs, this indirection would narrow sharply or disappear. As Chapter 9 explained, some CBDC models give citizens digital base money directly at the central bank, bypassing much of the commercial-bank buffer. From an intervention perspective, that shift is decisive because the state would need fewer third parties to compel. The closer the central bank moves toward being the bank, the more direct the account relationship becomes.

This changes the economics of intervention. In the strongest versions, observation no longer depends on subpoenas because the state reads its own ledger. Control no longer depends on the same regulatory mandates because the state administers its own system. The costs and frictions of triangular intervention shrink sharply. What once required legal process and institutional compliance can move much closer to automatic administration.

Total Intervention

Central Bank Digital Currencies can combine all three intervention types into a unified control mechanism. Some CBDC architectures support programmability and can expand surveillance and control well beyond what cash allows, while others are publicly framed as more privacy-preserving or more tightly intermediated.[15] Programmability is a design choice within digital currency, and a CBDC could be designed to replicate more cash-like properties: anonymous or bearer-like, and free of spending restrictions. Institutional incentive runs in one direction: most public proposals move toward greater traceability and control than physical cash allows.

CBDC rules can prohibit transactions directly through what amounts to autistic intervention embedded in the monetary infrastructure. The currency itself refuses to execute disfavored payments. Prohibitions on anonymous transactions or purchases of restricted goods, as well as payments to blacklisted recipients, become automatic. No prosecution is needed; the transaction fails.

Where CBDCs establish direct central bank accounts for citizens, they enable binary intervention without third-party collection: the state extracts transaction data directly, and economic activity be-

comes visible to the monetary authority at the point of settlement. There is no subpoena to issue and no bank to compel. Observation becomes much more immediate because the state is closer to being the counterparty to every balance.

A strongly programmable CBDC can also impose on private transactions through triangular intervention, for example by pressuring merchants to accept the currency, verify customer compliance status, or enforce spending restrictions. The state then controls exchanges it is not party to through the monetary medium itself.

Programmable Control

Certain CBDC designs can support interventions impossible with physical cash or current digital money. Money can be programmed to expire, forcing spending and preventing saving; this implements negative interest rates without the zero lower bound. It can be constrained to specified regions, so that a welfare payment might spend only at approved vendors or a regional stimulus might not leave the targeted area. It can refuse purchase categories, blocking payments for alcohol, tobacco, firearms, or politically disfavored goods at the monetary level. Transactions can require verified identity of both parties, making anonymous exchange impossible and ensuring every transaction is attributed. Some designs could also tie spending to broader compliance conditions.

The Two-Tier Illusion

Some central banks propose "two-tier" or intermediated CBDC designs that maintain commercial banks as customer-facing intermediaries, claiming this preserves the existing architecture. That claim needs qualification.

In a two-tier or intermediated CBDC, commercial banks or payment service providers remain customer-facing while the central bank retains issuance and core infrastructure functions. The commercial bank can become more of a front-end interface than an independent monetary intermediary, though the degree of disintermediation depends on the design. The buffer that Chapter 9

identified, where citizens hold claims on commercial banks and not direct central bank liabilities, may narrow sharply depending on how much control the central bank retains over settlement and rules.

Because the money is its liability, the central bank can set or constrain key rules around settlement and access. Depending on the legal and technical architecture, it may also gain broad visibility into retail payments or the ability to impose restrictions through intermediaries. Commercial banks in a two-tier CBDC can remain service providers operating within parameters the central bank defines instead of fully independent institutions creating money substitutes. The surveillance and control capabilities can remain substantial even when the user interface is delegated.

The Adoption Mechanism

Praxeological analysis reveals why explicit prohibition is unnecessary for CBDC adoption.

Autistic intervention (banning cash) is costly: it requires enforcement and generates resistance, while also making the control objective visible. The state has strong incentive to achieve the same outcome through less resistant means. If citizens adopt CBDCs voluntarily, enforcement costs vanish and political opposition never mobilizes.

The incentive structure favors voluntary adoption because CBDCs can offer real conveniences such as instant settlement, integration with government services, and smartphone accessibility. For individuals with high time preference, immediate convenience outweighs abstract future risks. Revealed preference in existing digital payment adoption shows that most individuals choose convenience over privacy when the tradeoff is not salient. The state need only ensure CBDCs are more convenient than alternatives.

Network effects create path dependency that is hard to reverse once established. As CBDC adoption increases, cash-handling infrastructure becomes less economical: banks face costs maintaining cash services for declining usage, while merchants face costs handling physical currency for fewer transactions. Each actor, pursuing their individual interest, rationally reduces cash infrastructure. The

aggregate effect is that cash becomes progressively less practical regardless of legal status.

This dynamic has historical precedent. Sweden's use of cash fell sharply over the 2010s without a formal cash ban, reaching roughly one in ten purchases as digital payments became dominant and cash infrastructure contracted.[16] The economic logic operates independently of CBDC specifics.

The likely endpoint of this path is clear enough even if the timeline is not: when cash infrastructure contracts beyond practical usability, CBDC usage can become mandatory in effect though voluntary in form. Surveillance that individuals first choose for convenience can become surveillance that is costly to escape without exiting the national currency entirely.

The Human Rights Foundation tracks this progression globally. As of 2025, more than 130 countries and currency unions are exploring CBDCs, though these activities differ substantially.[17] Research and study are widespread, but full deployment remains limited. Most jurisdictions are in exploratory phases, running pilots with limited participation or studying technical feasibility. Exploration does not commit a jurisdiction to deployment, and many pilot programs are discontinued before they reach the public. Several authoritarian regimes have launched retail CBDCs, with China's digital yuan the most prominent example, while democratic nations generally proceed more cautiously. The Foundation argues that autocracies remain the leading edge of live deployment.

The Rollout Record

Empirical results from the first wave of retail CBDC launches illuminate what the theoretical analysis predicted and what it did not.

The Nigerian eNaira is the clearest negative case. The Central Bank of Nigeria launched it in October 2021 as the first retail CBDC in a large African economy. Adoption remained below one percent of the adult population through 2024, and in late 2024 the central bank withdrew a substantial fraction of circulating tokens as part of a broader policy shift. Nigeria did not show that CBDCs fail. Nigeria

showed that the adoption mechanism described above depends on the CBDC being more convenient than the available alternatives, and Nigerians had access to mobile-money services that were already more convenient than the eNaira. When the state's product is less convenient than the incumbent and the autistic alternative of cash prohibition is too politically costly to pursue, the voluntary-adoption pathway fails.

China's e-CNY is the clearest positive case in a state that has both the regulatory capacity and the surveillance ambition to impose adoption. The People's Bank of China reports cumulative transaction volumes in the trillions of renminbi and deployment across every major city. Deployment depth tracks regulatory depth. Adoption is highest in jurisdictions where state employers and local tax authorities and merchant regulations together steer transaction flow into the system. China confirms what section 10.5 argued. A retail CBDC reaches deployment scale when the regulatory architecture around it routes payments into it and out of alternatives, and the monetary design choices matter less than the institutional pressure applied by the issuer.

The European Central Bank moved the Digital Euro from its preparation phase into its next phase at the end of October 2025. On the published schedule, the pilot begins in late 2026 assuming the supporting legislation is adopted, with potential first issuance in 2029. Design decisions remain open on the privacy architecture, on the holding limit, on the online-versus-offline treatment, and on the range of programmability permitted at the protocol level. The open questions illustrate that the politics of CBDC deployment in a democratic monetary union turn on the currency's capabilities, and the public debate in the member states has focused on whether an offline mode can deliver cash-like privacy in practice and whether programmability can be constrained by statute.

The United States moved in the opposite direction. Executive Order 14178, signed in January 2025, prohibits federal agencies from developing or issuing a U.S. retail CBDC. The order terminated the research activity authorized under the preceding administration. Regulated dollar-backed stablecoins are the explicit alternative fa-

vored by the order. A retail CBDC and a regulated dollar-backed stablecoin share many of the same surveillance capabilities when the stablecoin issuer is subject to KYC and AML and asset-freezing obligations. The difference is that the stablecoin leaves the surveillance compulsion with private issuers, which preserves the triangular-intervention mechanism the chapter has already analyzed, whereas the retail CBDC would have moved the compulsion directly onto the central bank. A preference for the older architecture over the newer one is not a rejection of financial surveillance.[18]

Cross-Border Settlement and the BRICS Split

The retail CBDC question is one axis. The wholesale cross-border question is a separate axis, and its politics have moved quickly enough that the history now divides cleanly.

The Bank for International Settlements' mBridge project connected the central banks of China, Hong Kong, Thailand, and the United Arab Emirates, with the Saudi Central Bank joining in mid-2024. The project's stated goal was to demonstrate multi-CBDC cross-border settlement without correspondent banking. In late 2024 the BIS withdrew from the project. The public framing emphasized that mBridge had reached minimum viable product status and that the BIS had completed its role. Several analyses published afterward observed that mBridge had become usable as an alternative to correspondent channels that run through the U.S. dollar and SWIFT, and that the withdrawal coincided with the political sensitivity of the BIS hosting such an alternative in the context of the 2024 BRICS summit. Whatever the causal weight assigned to those factors, the project has continued without BIS participation.

The BIS launched Project Agora in April 2024 as a parallel effort connecting the Federal Reserve Bank of New York, the European Central Bank, the Bank of England, the Bank of France, the Bank of Japan, the Bank of Korea, the Bank of Mexico, the Swiss National Bank, and the Central Bank of Brazil. Project Agora is designed around wholesale CBDC and tokenized commercial-bank deposits, and its participant list is the set of central banks in political alignment with the post-Bretton Woods financial order. The cross-border

monetary architecture has split along the same lines the surveillance architecture has split. One group of central banks is building cross-border rails that route around U.S.-denominated settlement and its associated reporting regime. Another group is building cross-border rails that integrate with it. The choice of rails determines the choice of observer.[19]

The Travel Rule and the Chain-Analysis Layer

Financial surveillance does not stop at the bank perimeter. The Financial Action Task Force's Recommendation 16, commonly called the Travel Rule, requires that transfers above a threshold carry originator and beneficiary information through the entire payment chain. Extended to virtual-asset service providers in 2019 and rolled out through 2024 across most jurisdictions that regulate cryptocurrency exchanges, the rule extends the triangular-intervention architecture into the crypto-asset market. The architecture is the same as the banking architecture: regulated intermediaries are compelled to surveil on the state's behalf, and the compulsion extends to every transfer that crosses an intermediary boundary.

Chain-analysis firms, including Chainalysis, Elliptic, and TRM Labs, sell the complementary capability. Their products ingest the public blockchain record and cluster addresses into entities using heuristics that the firms themselves treat as trade secrets. One KYC touchpoint anywhere in a cluster's history labels the cluster's real-world owner, and the label propagates across the cluster's subsequent transactions. The chain-analysis industry is the intelligence layer that converts blockchain transparency into identified enforcement targets. Chapter 20 examines the specific heuristics and their limits; the point here is that the chain-analysis layer is the enforcement infrastructure that makes the Travel Rule operational, and its existence illustrates why transparency is not a substitute for privacy in any regulated intermediary market.[20]

The Surveillance Endpoint

In their strongest surveillance-oriented forms, CBDCs are the culmination of financial surveillance. Every payment, to any party, for any amount, can be recorded and attributed; the full financial history envisioned by earlier surveillance programs moves much closer to automatic collection. Unlike post-hoc reporting, CBDC transactions are visible immediately, so intervention can occur before completion and not after detection. If CBDCs become mandatory and cash is eliminated, little or no unsurveilled economic activity remains inside the national-currency system; underground economies must develop alternative currencies entirely.

Sound money requirements from Chapter 9 therefore include censorship resistance and privacy. Money that can be surveilled and controlled fails as money in the Mengerian sense and functions instead as a mechanism of intervention.

10.6 The Adversarial Decision Cycle

State Surveillance Follows the OODA Pattern

Chapter 1 introduced John Boyd's OODA loop: Observe, Orient, Decide, Act.[21] Every intervention examined in this chapter follows this pattern.

Financial surveillance infrastructure exists to enable observation: CTRs report cash transactions, SARs flag suspicious activity, KYC requirements collect identity documents, third-party reporting creates transaction records. The Bank Secrecy Act and its successors are about observation. Without observation, subsequent stages cannot proceed.

Observation feeds orientation, where raw data becomes actionable intelligence through pattern correlation and suspicious-indicator evaluation, leading to target identification. KYC data becomes useful here: linking transactions to identified individuals enables the orientation that anonymous transactions would prevent. Orientation produces prioritized targets for decision, where finite resources are allocated among potential investigations. Finally, action takes the form of asset seizure, prosecution, regulatory sanction,

or account closure, but action is only possible against identified targets whose behavior has been observed and selected. The entire apparatus of enforcement depends on the preceding stages.

Breaking the Loop at Observation

Privacy technology breaks this cycle at its most vulnerable point. If the state cannot observe, it cannot orient on patterns it does not see, cannot decide to investigate transactions it does not know occurred, cannot act against targets it cannot identify.

Consider the contrast between traditional banking and carefully managed Bitcoin privacy. Traditional banking provides full observation: every transaction is logged with the account holder's identity attached and reported to authorities, and the state's OODA loop operates smoothly from observation through action. Bitcoin transactions remain public, but privacy practices examined in Chapter 20 can reduce easy attribution and degrade the state's observation quality. The decision cycle becomes harder to start cleanly, not magically blind.

States therefore work aggressively to restore observation. KYC requirements at exchanges attempt to link Bitcoin transactions to identities, restoring the observation that Bitcoin's pseudonymity impairs. Travel-rule regimes attempt to extend reporting requirements to many transfers handled by regulated crypto intermediaries. Blockchain analysis firms sell orientation capabilities, pattern-matching services that attempt to deanonymize transactions. Each effort aims to restore the observation that enables subsequent stages.

Cost Asymmetry

The OODA framework reveals a fundamental cost asymmetry. Observation infrastructure is expensive: building and maintaining the financial surveillance apparatus requires substantial ongoing investment in reporting systems, analysis capabilities, storage, and personnel. Each stage of the loop consumes resources.

Privacy, by contrast, can be cheap. Generating a cryptographic key costs nothing, and using CoinJoin[22] or Lightning Network[23]

adds minimal friction. Running transactions through privacy-preserving infrastructure imposes modest costs on the user while imposing massive costs on the adversary attempting to restore observation.

This asymmetry explains why privacy is strategic. The defender who prevents observation imposes costs far exceeding their own expenditure. The attacker who must overcome privacy protections faces costs that may exceed the value of the intelligence gained. When observation becomes expensive enough, the entire attack cycle becomes uneconomical.

CBDCs as Observation Infrastructure

Central Bank Digital Currencies attack the observation stage directly. Transactions moving through state-controlled infrastructure are visible close to the point of settlement, so the state no longer relies on compelled third-party reporting or incomplete data. By sharply narrowing the possibility of unobserved transactions, a retail CBDC would attack the most effective point at which privacy breaks the decision cycle. The response developed in Part V is to build monetary systems where observation is structurally expensive. The goal is not perfect unobservability but raising observation costs beyond what adversaries are willing to pay.[24]

Chapter Summary

Financial surveillance operates through Rothbard's intervention typology. Autistic intervention prohibits privacy tools directly; binary intervention extracts information from individuals through subpoena, seizure, and compelled disclosure; triangular intervention, the dominant form, forces private institutions to surveil on the state's behalf. The Bank Secrecy Act and its successors built the third mechanism into the core of the financial system, and the Misesian logic of intervention explains why the apparatus has expanded across every decade since. Central Bank Digital Currencies, in their strongest surveillance-oriented forms, narrow or remove the commercial-bank buffer between state and citizen and can combine

all three intervention types into a single programmable instrument. The FATF Travel Rule and the chain-analysis industry extend the same triangular logic to virtual-asset service providers, converting blockchain transparency into identified enforcement targets by anchoring address clusters to KYC touchpoints.

Nigeria's eNaira stayed below one percent adoption because voluntary adoption requires the CBDC to outcompete incumbents, and mobile money was already more convenient. China's e-CNY reached every major city through regulatory pressure routing transaction flow into the system. The EU's Digital Euro is in deliberative preparation for potential 2029 issuance, and the U.S. reversed course through Executive Order 14178 (January 2025), prohibiting federal retail CBDC development and favoring regulated dollar-backed stablecoins instead. Cross-border settlement split along the same lines: the BIS withdrew from mBridge in late 2024, leaving it as a BRICS-aligned alternative, while Project Agora, launched April 2024, carries the Western-aligned parallel. The architecture splits by political alignment, and the choice of rails determines the choice of observer.

Privacy breaks the state's decision cycle at observation. Boyd's OODA loop depends on it, and when privacy technology degrades observation, orientation, decision, and action all degrade behind it. The cost asymmetry favors defenders: cryptographic defense is cheap and attack is expensive. This does not make every CBDC design equally surveillance-oriented; the cash-like bearer designs sit at one end of the spectrum and the fully programmable systems at the other, and it does not make the state blind. Chapter 20 examines the specific tools and their limitations; this chapter has set the threat model, and Part V develops the technical response.[25]

Notes

[1]Franz Oppenheimer, *The State*, trans. John M. Gitterman (Indianapolis: Bobbs-Merrill, 1914), 15.

[2]Murray N. Rothbard, *Power and Market: Government and the Economy*, 4th ed. (Auburn, AL: Ludwig von Mises Institute, 2006), chapters 1–2.

[3]U.S. export controls on cryptographic software were codified in the Export Administration Regulations (EAR) under the International Traffic in Arms Regulations (ITAR) and later EAR Category 5, Part 2 (Information Security), treating strong encryption as a munition requiring State Department export license through 1996. The Clinton administration's relaxation followed litigation in *Bernstein v. United States Department of Justice*, 922 F. Supp. 1426 (N.D. Cal. 1996), in which Judge Marilyn Patel held that source code constitutes protected speech under the First Amendment. Export Administration Regulations, 15 C.F.R. Parts 730–774. On the Clipper Chip and key-escrow proposals (1993–1996), see Steven Levy, *Crypto: How the Code Rebels Beat the Government — Saving Privacy in the Digital Age* (New York: Viking, 2001), chapters 10–13. Countries retaining encryption restrictions or mandatory key escrow include Russia (FSB registration requirements under SORM) and China (commercial encryption licensing under the Cryptography Law of 2020).

[4]"Coinjoining" refers to the CoinJoin protocol, a privacy technique for Bitcoin transactions. For the original technical proposal, see Gregory Maxwell, "CoinJoin: Bitcoin Privacy for the Real World," BitcoinTalk forum post (August 22, 2013), `https://bitcointalk.org/index.php?topic=279249.0`. For the regulatory context around coinjoining services, see FinCEN guidance on mixing services and the prosecutions of Tornado Cash and Samourai Wallet developers discussed in Chapter 13.

[5]Civil asset forfeiture in U.S. federal law proceeds under 18 U.S.C. § 981 (civil forfeiture) and 21 U.S.C. § 881 (drug-related forfeiture), with state equivalents across all fifty states. The constitutional standard requires only probable cause that the property is connected to crime; the owner bears the burden of proving innocent ownership. Leading cases include *United States v. $8,850 in United States Currency*, 461 U.S. 555 (1983), and *Bennis v. Michigan*, 516 U.S. 442 (1996). For systematic documentation of abuses, see the Institute for Justice, *Policing for Profit: The Abuse of Civil Asset Forfeiture*, 3rd ed. (2020), `https://ij.org/report/policing-for-profit/`. In 2019 Congress required that federal agencies report forfeiture proceeds to FinCEN, creating the data trail that documents scale; for the fiscal data, see U.S. Department of Justice, *Annual Asset Forfeiture Report*, `https://www.justice.gov/afp/annual-reports`.

[6]The third-party doctrine holds that information voluntarily shared with a third party loses Fourth Amendment protection. The foundational cases are *United States v. Miller*, 425 U.S. 435 (1976) (bank records) and *Smith v. Maryland*, 442 U.S. 735 (1979) (phone numbers dialed). The Supreme Court partially limited the doctrine in *Carpenter v. United States*, 585 U.S. 296 (2018), holding that cell-site

location information collected over seven days requires a warrant, but the core doctrine remains intact for financial records. On the tension between the doctrine and digital-era surveillance, see Paul Ohm, "The Fourth Amendment in a World Without Privacy," *Mississippi Law Journal* 81 (2012): 1309; and Orin S. Kerr, "The Case for the Third-Party Doctrine," *Michigan Law Review* 107 (2009): 561.

[7]The Axiom of Resistance is Eric Voskuil's formulation; see Chapter 5, note 1, for the primary citation.

[8]For the evolution of BSA reporting requirements, see U.S. Department of the Treasury, Financial Crimes Enforcement Network, "History of Anti-Money Laundering Laws," available at `https://www.fincen.gov/history-anti-money-laundering-laws`.

[9]On the currency monopoly, seigniorage, and the praxeological analysis of the Federal Reserve's exclusive money-creation privilege, see Murray N. Rothbard, *The Mystery of Banking*, 2nd ed. (Auburn, AL: Ludwig von Mises Institute, 2008 [1983]), `https://mises.org/library/mystery-banking`, especially chapters 1–4 on money and banking and chapters 10–14 on central banking. For Mises's treatment of the relationship between monopoly currency issuance and inflation as a form of taxation, see *Human Action*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 1998 [1949]), chapter 17 ("Indirect Exchange"), sections 8–12. For the legal structure of the Federal Reserve's issuance authority, see Federal Reserve Act of 1913, 12 U.S.C. § 411 (issuance of Federal Reserve notes as legal tender).

[10]Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008), `https://bitcoin.org/bitcoin.pdf`; reference implementation at Bitcoin Core, `https://bitcoincore.org/`.

[11]Rothbard analyzes intervention cascade effects in *Power and Market*, chapter 3; see also *Man, Economy, and State*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 2009), chapter 13. Mises's underlying treatment is in *Human Action: A Treatise on Economics*, Scholar's Edition (Auburn, AL: Ludwig von Mises Institute, 1998 [1949]), Part Six, especially chapter 30 ("Interference with the Structure of Prices"); and *Interventionism: An Economic Analysis*, ed. Bettina Bien Greaves (Indianapolis: Liberty Fund, 2011 [1940]). The argument that each intervention invites the next, with middle-of-the-road policies producing either a slide into socialism or retreat toward the free market, is the central structural claim of both Mises works.

[12]Money Laundering Control Act of 1986, Pub. L. 99-570, 100 Stat. 3207 (1986), codified at 18 U.S.C. §§ 1956–1957. The Act created the first federal money-laundering offense as a standalone crime (not merely a predicate of drug trafficking) and required financial institutions to file Suspicious Activity Reports. It was the first major legislative extension of the Bank Secrecy Act framework.

[13]Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act (USA PATRIOT Act) of 2001, Pub. L. 107-56, 115 Stat. 272 (2001). Title III ("International Money Laundering Abatement and Anti-Terrorist Financing Act") extended BSA obligations to money-service businesses, broker-dealers, and other non-bank financial

institutions, closing regulatory gaps that had allowed alternative remittance channels to operate outside the reporting regime. For FinCEN's implementing regulations, see 31 C.F.R. Part 1022 (money-service businesses).

[14]FinCEN Customer Due Diligence Requirements for Financial Institutions, 81 Fed. Reg. 29398 (May 11, 2016), codified at 31 C.F.R. § 1010.230. The rule added a fifth pillar to the BSA/AML framework: covered financial institutions must identify and verify the beneficial owners of legal-entity customers (natural persons owning 25% or more of equity, plus one control-person nominee), closing the shell-company gap the prior regime left open. Amended by the Corporate Transparency Act of 2021, Pub. L. 117-54, which moved beneficial-ownership reporting to FinCEN's own registry rather than relying solely on bank-level collection.

[15]For economic analysis of CBDCs, see Jörg Guido Hülsmann, *The Ethics of Money Production* (Auburn, AL: Ludwig von Mises Institute, 2008). For technical analysis, see Agustín Carstens (General Manager, BIS), "Digital Currencies and the Future of the Monetary System," speech at the Hoover Institution, January 2021.

[16]Sveriges Riksbank, *Payments Report 2023*, `https://www.riksbank.se/`, documents the long-run decline of cash in Sweden across the 2010s and into the early 2020s; the share of payments made in cash at point of sale fell from roughly 39 percent in 2010 to roughly 9 percent by the end of the decade.

[17]Human Rights Foundation, "CBDC Tracker," available at `https://cbdctracker.hrf.org`. The tracker monitors CBDC development globally with particular attention to civil liberties implications. Launched in 2023, it is the first monitoring effort focused specifically on the human rights impact of central bank digital currencies. The Foundation notes that dictatorships are leading CBDC deployment, with nearly half the world's population living under autocracies actively developing these systems.

[18]The 2021–2025 retail CBDC rollout record divides into a negative case (Nigeria), a positive case (China), a democratic-deliberation case (EU), and a reversal case (U.S. Executive Order 14178 in January 2025 prohibiting federal retail CBDC development). Central Bank of Nigeria eNaira, `https://enaira.gov.ng/`. People's Bank of China digital yuan (e-CNY), overview at `http://www.pbc.gov.cn/en/3688006/3688253/4157443/index.html` and test-app information at `https://money.163.com/special/ecny/`. European Central Bank Digital Euro project, `https://www.ecb.europa.eu/paym/digital_euro/html/index.en.html`. U.S. Executive Order 14178, "Strengthening American Leadership in Digital Financial Technology," 90 Fed. Reg. 8647 (January 23, 2025), `https://www.federalregister.gov/documents/2025/01/31/2025-02123/strengthening-american-leadership-in-digital-financial-technology`. Cross-jurisdictional tracker at Atlantic Council, "Central Bank Digital Currency Tracker," `https://www.atlanticcouncil.org/cbdctracker/`, and Human Rights Foundation, `https://cbdctracker.hrf.org/`.

[19]The BIS withdrew from mBridge in late 2024, and mBridge now runs as a BRICS-aligned cross-border settlement alternative; Project Agora, launched April 2024, is the BIS's parallel effort with Western-aligned central banks; the two projects split cross-border settlement rails along the same lines the surveillance architecture has split. BIS mBridge project page, https://www.bis.org/about/bisih/topics/cbdc/mcbdc_bridge.htm. BIS announcement of withdrawal, October 31, 2024, https://www.bis.org/press/p241031.htm. BIS Project Agora page, https://www.bis.org/about/bisih/topics/fmis/agora.htm, with April 2024 launch announcement at https://www.bis.org/press/p240403.htm. Independent analysis of the geopolitical context in Atlantic Council and CSIS commentary, accessible at https://www.atlanticcouncil.org/ and https://www.csis.org/.

[20]The FATF Travel Rule (Recommendation 16, extended to virtual-asset service providers in 2019) requires regulated crypto intermediaries to transmit originator and beneficiary information across transfers; Chainalysis, Elliptic, and TRM Labs are the chain-analysis firms that convert blockchain transparency into identified enforcement targets. FATF Recommendation 16 (Travel Rule), "Updated Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers" (October 2021), https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Guidance-rba-virtual-assets-2021.html. Chainalysis, https://www.chainalysis.com/. Elliptic, https://www.elliptic.co/. TRM Labs, https://www.trmlabs.com/. On Travel Rule implementation across jurisdictions, see the FATF mutual-evaluation reports at https://www.fatf-gafi.org/en/publications/Mutualevaluations.html. Critical analysis of chain-analysis heuristics in Sergio Demian Lerner and others' blog posts at https://bitslog.com/ and Coinbase's public analyses of chain-analysis methodology.

[21]For Boyd's primary OODA-loop formulation, see Chapter 1, note 4.

[22]CoinJoin is the primary Bitcoin privacy technique combining multiple users' inputs and outputs into a single transaction, making it harder to trace individual payment flows. For the original protocol description, see Gregory Maxwell, "CoinJoin: Bitcoin Privacy for the Real World," BitcoinTalk (August 2013), https://bitcointalk.org/index.php?topic=279249.0. Major implementations include Wasabi Wallet (WabiSabi protocol, https://wasabiwallet.io/) and JoinMarket (https://github.com/JoinMarket-Org/joinmarket-clientserver). For analysis of CoinJoin's privacy guarantees and limitations, see Adam Ficsor (nopara73) et al., "WabiSabi: Centrally Coordinated CoinJoins with Variable Amounts," Financial Cryptography and Data Security 2022; and Malte Möser et al., "An Empirical Analysis of Traceability in the Monero Blockchain," *Proceedings on Privacy Enhancing Technologies* 3 (2018), which contextualizes CoinJoin by comparison.

[23]The Lightning Network is a second-layer payment protocol for Bitcoin enabling fast, low-cost payments through off-chain payment channels that settle on the base layer. The original specification is Joseph Poon and Thaddeus Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments" (January 14, 2016), `https://lightning.network/lightning-network-paper.pdf`. The protocol is specified in the BOLT (Basis of Lightning Technology) standards at `https://github.com/lightning/bolts`. For privacy properties and limitations, see Rene Pickhardt and Mariusz Nowostawski, "Imbalance Measure and Risk-Minimization for Lightning Network Payment Channel Rebalancing," and George Kappos et al., "An Empirical Analysis of Privacy in the Lightning Network," *Financial Cryptography and Data Security* (2021), which analyzes probing attacks and routing-graph deanonymization.

[24]Further reading on CBDCs and financial surveillance: see note 25 below.

[25]Further reading on CBDCs and financial surveillance. For the specific CBDC design space, the Bank for International Settlements' working papers are the broadest official source: Raphael Auer, Giulio Cornelli, and Jon Frost, "Rise of the Central Bank Digital Currencies: Drivers, Approaches and Technologies," BIS Working Paper No. 880 (2020), `https://www.bis.org/publ/work880.pdf`; Codruta Boar and Andreas Wehrli, "Ready, Steady, Go? Results of the Third BIS Survey on Central Bank Digital Currency," BIS Papers No. 114 (2021); and the annually updated CBDC surveys cited in the preface. On the political economy, Eswar S. Prasad, *The Future of Money: How the Digital Revolution Is Transforming Currencies and Finance* (Harvard University Press, 2021), is the balanced academic treatment; Christopher Leonard, *The Lords of Easy Money: How the Federal Reserve Broke the American Economy* (Simon & Schuster, 2022), supplies the institutional context. For the civil-liberties case against retail CBDCs, Nicholas Anthony and Norbert Michel, "Central Bank Digital Currency: Assessing the Risks and Dispelling the Myths," Cato Institute Policy Analysis No. 941 (2023), `https://www.cato.org/policy-analysis/central-bank-digital-currency`; and Human Rights Foundation, "The Case Against Central Bank Digital Currencies," at `https://hrf.org`. On China's digital yuan (e-CNY) as the leading deployed example, Eswar Prasad's work again; for the EU's digital-euro trajectory, see ECB publications at `https://www.ecb.europa.eu/paym/digital_euro/html/index.en.html`. For the broader financial-surveillance literature: Nicholas Anthony, *The Infrastructure of Control: A Global Look at the Dangers of CBDCs* (Cato, 2024); Allen J. Dickerson and Lloyd M. Pierson, "Central Bank Digital Currencies: Privacy Implications," *Journal of Law and Liberty* (2022). On the underlying legibility problem the state solves through monetary surveillance, James C. Scott's *Seeing Like a State* (Yale, 1998), already cited in Chapter 1, is the historical frame; Rainer Böhme et al., "Bitcoin: Economics, Technology, and Governance," *Journal of Economic Perspectives* 29, no. 2 (2015): 213–238, is the reference cryptocurrency-vs-state treatment. On the regulatory machinery specifically, Todd Zywicki and Max Raskin, "CBDCs in

the United States: A Constitutional Analysis" (2024), addresses the separation-of-powers concerns that U.S. retail CBDC deployment would raise.

11

CORPORATE SURVEILLANCE AND DATA EXTRACTION

> "Once a technology is admitted, it plays out its hand; it does what it is designed to do."
>
> Neil Postman, *Technopoly* (1992)[1]

States are not the only surveillance threat. Corporate data extraction has built monitoring infrastructure that rivals and often exceeds state capabilities, and its business model, entanglement with state surveillance, and relationship to market failure all bear on how markets are beginning to respond to privacy demand.

11.1 The Business Model of Data Extraction

Users as Product

Shoshana Zuboff coined the term "surveillance capitalism" to describe a business model where human experience is claimed as free raw material for translation into behavioral data.[2] This data feeds prediction products traded in behavioral futures markets.

The economic logic is simple: services appear free because users pay with data, not money. The advertiser is the customer; the user is the product. More precisely, predictions about user behavior are

the product, while users supply the raw material for manufacturing those predictions.

This inverts the traditional market relationship. In ordinary exchange, businesses compete to serve customers. In data extraction, businesses compete to capture users. Serving customers requires satisfying their preferences; capturing users requires keeping them engaged regardless of whether engagement serves their interests.

Attention Harvesting

Data extraction businesses compete for attention. User engagement generates data; more engagement generates more data; maximizing engagement maximizes the raw material for prediction products.

This creates incentives for manipulation. If engagement serves user interests, no problem arises. But when engagement conflicts with user interests (addictive design, outrage amplification, rabbit holes of increasingly extreme content), the business model rewards manipulation over service.

Praxeology emphasizes revealed preference: action reveals preference through what people choose. But interpretation becomes harder when choice architecture is built to exploit psychological vulnerabilities. The user who spends hours scrolling may be revealing some preference, but not necessarily for scrolling as such; they may also be revealing susceptibility to variable reward schedules.[3]

Behavioral Surplus

Zuboff distinguishes between data necessary for service improvement and "behavioral surplus" extracted for prediction products. A map application needs location data to provide directions; that same location data, accumulated over time and correlated with other data, becomes raw material for predicting future behavior and selling those predictions to advertisers.

The surplus concept highlights that users receive services worth some fraction of the data they provide. The remainder, the surplus, is extracted without compensation. Users cannot easily assess how

much surplus is extracted because they cannot observe how their data is used, combined, or sold.

This information asymmetry compounds the problem. In ordinary markets, competition drives prices toward marginal cost. In data extraction markets, users cannot comparison shop based on data extraction because they cannot observe extraction practices. Competition therefore occurs on other dimensions (features, network effects), not privacy protection.[4]

A Cantillon Analogy for Data

Richard Cantillon's eighteenth-century analysis of monetary flows observed that new money enters the economy at specific points and that the recipients closest to those entry points benefit before prices have adjusted to the expansion.[5] The Austrian tradition uses the concept narrowly, as a claim about the distributional effects of monetary expansion, and the claim holds up under that restriction.

A careful analogy runs in the data extraction economy without claiming the analogy is a proof. Behavioral records enter the commercial system at identifiable points (the browser, the phone, the smart device, the application session) and the firms closest to those collection points can combine and resell the records before the subject has any opportunity to react. Prices in the attention market adjust continuously as prediction accuracy improves, and the adjustment benefits the firms earliest in the chain. The similarity is structural, not monetary: first-access advantage produces disproportionate gains regardless of whether the substrate is currency or behavior.

Cantillon's argument about money rests on the specific fact that currency is scarce and that new issuance dilutes the purchasing power of existing holders, which has no direct analog in data flows. The analogy establishes only that the economic form of the pattern (first recipients gain before the rest of the system adjusts) recurs in any market whose raw material flows through identifiable entry points.

Prediction Products

Data extraction businesses sell predictions. Advertisers pay for likely-to-click users; political campaigns pay for likely-to-persuade voters; insurers pay for likely-to-claim customers. The value lies in prediction accuracy; accuracy improves with more data; more data requires more extensive surveillance.

This creates an extraction ratchet. Each improvement in prediction accuracy makes data more valuable, which justifies more intensive extraction and then enables more accurate predictions. The endpoint, approached asymptotically, is ever broader monitoring in support of ever more refined behavioral prediction.

11.2 Corporate-State Entanglement

Legal Requirements

Chapter 10 examined triangular intervention, which is state mandates imposed on private transactions, and data extraction businesses face such mandates in several forms. Data retention requirements force companies to keep data they might otherwise delete; lawful interception requirements force communication providers to build surveillance backdoors. Reporting requirements force platforms to monitor for specified content.

These requirements shape corporate data practices. A company might prefer to minimize data collection for security and liability reasons, but legal requirements may mandate collection. The state uses corporate infrastructure as force multiplier, achieving surveillance scope far harder to assemble through direct government operation alone.

Voluntary Cooperation

Beyond legal requirements, many corporations cooperate voluntarily with government requests. The PRISM program revealed extensive government access to data held by major technology companies.[6] National Security Letters compel disclosure while prohibiting recipients from acknowledging the request.[7]

Voluntary cooperation creates business opportunities. Government contracts reward companies with surveillance capabilities. Intelligence agencies are well-funded customers for prediction products. The line between serving advertisers and serving intelligence agencies becomes blurred when both want the same behavioral predictions.

The Public-Private Partnership

State and corporate surveillance have become symbiotic. States benefit from corporate data collection that would face legal barriers if conducted directly. Corporations benefit from legal frameworks that entrench their business models while burdening competitors.

Corporations collect data at scale that governments could not legally mandate, and governments then access that data through legal process, national security letters, or informal cooperation. Commercial analytical tools (machine learning, pattern recognition) serve the same surveillance ends as government programs. Corporate networks, devices, and platforms become the infrastructure on which monitoring rides. And large corporations can afford compliance with complex privacy regulations while smaller competitors cannot, so rules intended to protect privacy instead create moats protecting incumbents.

The Distinction Matters Less Than It Appears

The state-corporate distinction is consequential for legal purposes. Constitutional constraints apply to government action, not private action. For privacy analysis, the distinction is thinner.

If your communications are monitored, one core privacy loss can be similar whether the monitor is NSA or Google. If your behavior is predicted and manipulated, the injury does not disappear when the manipulator is a social media algorithm and not a government propaganda agency. The identity of the surveilling party still matters for law and downstream power, but the loss of privacy comes first.

This suggests that privacy protection must address both state and corporate surveillance. Technical measures effective against one may be effective against both. But legal measures effective against government surveillance (constitutional constraints, warrant requirements) do not reach corporate surveillance directly.

11.3 Is This a Market Failure?

The Market Failure Claim

Critics argue that surveillance capitalism is market failure.[8] Users do not want full surveillance but get it anyway. Companies extract negative externalities, like privacy costs, without bearing them. Markets fail to produce privacy-respecting alternatives.

If true, this would justify intervention to correct the failure. Privacy regulations, data ownership rights, platform breakups might be warranted to restore functioning markets.

The Austrian Response

Austrian economics is skeptical of market-failure claims on several grounds that apply together. The knowledge problem makes central assessment of the "correct" level of privacy impossible: optimal outcomes emerge through market process and cannot be known in advance, and what looks like failure may be markets discovering solutions to problems regulators have not yet identified. The current market structure also reflects decades of intervention that distort the baseline against which failure is being diagnosed; the relevant question is whether observed outcomes result from markets or from the interventions shaping them. And static analysis may identify apparent inefficiencies that dynamic analysis reveals as temporary, meaning markets may be in the process of correcting the problem while intervention arrests that correction.

The intervention record on surveillance specifically is substantial and easy to document. Copyright and patent law help entrench incumbents by raising the cost of copying protocols and building compatible software. GDPR[9] and similar regulations impose com-

pliance costs that large firms absorb more easily than smaller rivals, reducing some abusive practices while hardening incumbent advantage at the same time. Major technology companies derive substantial revenue from government contracts, which creates a direct incentive to build surveillance capabilities that states want to buy. Section 230 and similar provisions lower the legal risks of hosting and large-scale curation, which favors platform scale.[10] Banking regulations requiring know-your-customer and anti-money-laundering compliance push economic activity toward tracked digital channels and away from private cash, producing more of the data streams that surveillance businesses monetize. The "market" in which surveillance capitalism operates is far from a counterfactual free market. Its structure has been shaped at every layer by state policy, and the outcomes the critics attribute to the market partly reflect those policy choices.

Would a Free Market Produce This?

The counterfactual is difficult to assess, but several considerations suggest current outcomes are not inevitable market results. Without IP protection creating artificial scarcity in software, competition would be more intense; platforms could not maintain network effects by threatening compatible alternatives with patent lawsuits. If users could easily switch between compatible platforms, data extraction would face competitive pressure, and users who value privacy could migrate to privacy-respecting alternatives without losing network connections. Without monetary system surveillance pushing transactions toward tracked channels, alternative payment methods would reduce the data streams that make behavioral prediction valuable.

This does not prove that free markets would produce perfect privacy. Network effects, coordination problems, and real consumer preferences for free services would still exist. But it suggests current surveillance intensity reflects intervention as much as market outcome.

Network Effects and Lock-In

Even granting that intervention plays a role, real market dynamics contribute to surveillance concentration. Communication platforms are more valuable with more users, creating winner-take-all dynamics where a few platforms dominate and making exit costly for users. Users have invested in building connections and creating content on these platforms; switching means abandoning those investments. Even if users prefer privacy-respecting alternatives, coordination failure may prevent migration, as each user waits for others to switch and no one wants to be first to an empty platform.

These are real market phenomena, not intervention effects. They suggest that even absent intervention, privacy competition might face obstacles. But they do not justify further intervention; they suggest the problem requires technical and entrepreneurial solutions, not regulatory ones.

The Seen and the Unseen

Frédéric Bastiat's 1850 essay on the seen and the unseen is the correct frame for the cost accounting that "nothing to hide" discussions typically omit.[11] The good economist, Bastiat wrote, considers not only the immediate effect of an action but the effects it produces in the longer term and among the groups other than the one directly in view. Henry Hazlitt made the same principle the organizing lesson of *Economics in One Lesson* (1946), and the principle applies here without modification.

The seen effects of commercial surveillance are easy to list: the free service, the personalized feed, the convenient autofill, the advertisement that reaches the advertiser's intended audience. The unseen effects are the ones that require the economist's discipline to notice. An idea not shared because the author discounted the probability that a future employer would read the post. An entrepreneurial attempt not made because the record of its failure would be permanent and attached to the entrepreneur's legal name. A purchase not made because the transaction would have revealed the purchaser's interest in a disfavored topic. A consultation not

scheduled because the intake form asks questions the patient expects to see used against them later. Each case is a specific voluntary exchange that surveillance taxed out of existence, and the aggregate is the economy the surveillance infrastructure has suppressed.

The unseen effects are larger than the seen, because most of what surveillance suppresses never happens at all. The free service is visible on every screen. What did not happen produces no data, and its cost cannot be measured from the outside. Analysis can still name the cost, and the naming is the discipline Bastiat insisted on.

11.4 The Irreversibility Problem

The analysis so far has treated exposed data as a category. That treatment is incomplete. One class of data has a property the others do not, and the property forces a distinct defensive posture. Biometric and genetic data cannot be rotated. Once exposed, the exposure is permanent.

Why Rotation Is the Usual Remedy

Credential exposure has a standard remedy. A leaked password is replaced with a new one; a compromised key is revoked and superseded by a new key pair; a stolen credit card number is cancelled and reissued; a leaked session token expires on its own schedule. The remedy works because the token is an artifact that can be replaced without changing the person it represents. Credential rotation is the defensive posture the security literature assumes whenever it describes "compromise" as a condition with a recovery path.

The same logic applies to most behavioral and financial data, though imperfectly. Exposed transaction records cannot be retracted, but the adversary's knowledge of them ages: the pattern of purchases made two years ago describes a person who may have moved, changed jobs, and changed tastes. The information still has value to a chain-analysis firm or a data broker, but its value decays. Rotation-by-aging is slower than rotation-by-key-replacement, and it depends on the subject living differently over time, but it preserves the principle that exposure is a condition with a recovery path.

Why Biometric and Genetic Data Are Different

Biometric and genetic data are the defining category for which neither rotation mechanism works. A fingerprint is the same fingerprint in 2025 that it was in 2005 and that it will be in 2045. A face is the same face; an iris is the same iris; a voice print drifts only modestly with age. A genome is identical across every cell of a person's body and stable across every year of a person's life. The identifier is the person. The person cannot be reissued.

The practical consequence is that a database breach of biometric or genetic data is not a recoverable condition. The breach of a credential database means "rotate all credentials now"; the breach of a biometric database means "the identifier is now permanently known to the adversary, and every future use of that identifier must be assumed to be observable by the adversary." The exposure extends across the lifetime of the person.

The insurance analogy is telling. Standard cyber-insurance policies price credential breaches using formulas that assume remediation ends when the credentials are rotated. Formulas applied to biometric breaches understate the true cost by orders of magnitude, because the exposure does not end. Several underwriters have recognized the discrepancy and adjusted their pricing; the adjustment is a market-side recognition that biometric and genetic data belong to a different category of asset.

The Cases Illustrate the Principle

Two cases illustrate the principle concretely enough to anchor the analysis. The first is the breach of a major consumer genetic-testing company in late 2023, in which the profiles of millions of customers were exposed through credential-stuffing attacks and then advertised for sale on underground forums segmented by ethnicity. The company subsequently entered bankruptcy and the genetic database became a primary asset in the proceedings, changing hands to a successor organization. The customers who contributed their genomes had no role in that transfer, and no remedy could undo it.[12]

The second is the national biometric identity program of a state with more than a billion enrolled identities, in which fingerprints and iris scans and demographic data have been centrally held for years. The program has survived multiple alleged leaks. Independent of any specific allegation, a system of this kind creates a single point whose compromise would permanently enumerate an entire population's biometric identifiers, and its defensive posture cannot be rotation-based for the same reason as the first case.[13]

Chapter 12 examines a third pattern in detail: commercial facial-recognition corpora built from web-scraped images, where subjects never consented and regulatory fines have been absorbed as a cost of business. The structural property is the same.

The Defensive Posture

The defensive posture that follows from irreversibility is different from the posture that applies to credentials. Minimize enrollment in any biometric or genetic collection whose purpose does not justify lifetime exposure. Where enrollment is unavoidable (border control, a state identity program, a medical necessity), prefer systems designed with biometric cancelability, where the stored template is a one-way function of the biometric and a secret that can be rotated. Where the service is optional (a phone's face-based authentication, a consumer genetic-testing service, a biometric gym-membership system), weigh the ordinary convenience of the enrollment against its irreversible exposure. For most consumers under most threat models, the weighing does not justify the enrollment.

Legal reform has a narrow role here. State and national-level biometric privacy laws can create rights of action, deletion rights, and civil penalties; the laws of this class that have passed in several U.S. states since 2021 establish that the legal system can price biometric exposure after the fact. They cannot prevent the exposure in the first place. Prevention remains an engineering and personal-discipline problem, and the discipline must be calibrated to the irreversibility of the asset.[14]

The broader analytical point is that the taxonomy of surveillance must treat irreversible data as a distinct category. The earlier chap-

ters of the book analyzed financial surveillance, corporate surveillance, and the Crypto Wars as separable channels. Biometric and genetic data cut across all three channels and impose a common constraint: whatever else the defender does, the identifier is the person, and any defensive architecture that relies on rotating the identifier fails by construction. This is the category that forces the strictest defensive discipline, and the rest of the book's operational chapters take the discipline as given.

11.5 Market Responses and Privacy Competition

Privacy as Competitive Differentiator

Despite obstacles, markets have begun responding to privacy demand. Some established companies now differentiate on privacy; new entrants build privacy-first business models; infrastructure changes constrain data extraction regardless of individual company choices.

Apple's privacy differentiation provides the clearest large-scale demonstration. In April 2021, Apple introduced App Tracking Transparency (ATT), requiring apps to request permission before tracking users across other companies' apps and websites. The result was dramatic: approximately 80% of iOS users opted out of tracking when given a clear choice.[15] Meta reported that ATT would reduce its 2022 revenue by approximately $10 billion; industry estimates placed the total cost to Meta closer to $13 billion annually.[16]

This single policy change revealed the fragility of surveillance-dependent business models. When users were given a simple choice, the vast majority chose not to be tracked. A large underlying privacy preference was already there; ATT gave it a simple mechanism for expression. Apple profited from revealing it; companies dependent on surveillance suffered. This is market discovery operating at scale.

Search and browser alternatives show similar dynamics. DuckDuckGo[17] has grown from a niche search engine to processing billions of queries annually, despite competing against the most sophisticated search infrastructure in history. Users accept less

sophisticated results in exchange for privacy; the trade-off reveals a real willingness to pay. Brave browser[18] has reached tens of millions of users by combining privacy protection with attention-based advertising that compensates users instead of extracting from them.

The Rise of Encrypted Messaging

End-to-end encrypted messaging has achieved mainstream adoption more completely than any other privacy technology.

Signal's growth trajectory illustrates the market discovery process. In January 2021, following WhatsApp's announcement that it would share more data with Facebook, Signal's servers crashed under the load of new users.[19] The episode showed how quickly users will move when privacy costs become salient and a usable alternative exists. Signal's nonprofit structure also shows that encrypted messaging need not depend on advertising or data extraction.

WhatsApp itself, despite Meta ownership, uses the Signal protocol[20] for end-to-end encryption. The decision was defensive: without encryption, WhatsApp would lose users to encrypted alternatives. Even surveillance-dependent companies often need to provide some privacy features to remain competitive. This is market pressure operating through competition, not through regulation.

The encryption adoption pattern reveals something about how markets discover privacy demand. Encryption was once an expert-only technology that required manual key exchange and careful configuration, until Signal made it default and invisible and users benefit without needing to understand the technology. The lesson: privacy tools must be as convenient as surveillance alternatives to achieve broad adoption. Usability, not just security, strongly shapes market success.

Paid vs. Ad-Supported Models

The "free" services model depends on data extraction for revenue. Paid models offer an alternative: users pay with money,

not data. When users are customers and not products, business incentives align with user interests instead of against them.

Subscription services have grown across categories. Streaming video offers ad-free tiers; users revealed preference for paying to avoid surveillance-enabling ads. News paywalls remove the advertising incentive to maximize engagement regardless of content quality. Productivity software subscriptions have displaced advertising-supported tools, changing incentive structures across the software industry.

Not all subscription services respect privacy; paid products can still extract data. But the paid model removes the structural incentive that makes data extraction the core business instead of an incidental practice. A company whose revenue comes from subscriptions has no structural reason to maximize data collection; a company whose revenue comes from prediction products has every reason.

The premium tier pattern, appearing across products and services, suggests growing willingness to pay for privacy and reduced surveillance. Users who once accepted "free" services now pay for alternatives that better align with their interests. This revealed preference guides entrepreneurial discovery of further privacy-respecting products.

Privacy Infrastructure

Beyond individual products, infrastructure changes are beginning to constrain data extraction structurally. DNS-over-HTTPS[21] prevents ISPs from observing and monetizing browsing data. Default encryption in transit, now standard across the web, prevents casual interception. Hardware security modules in consumer devices make certain types of data extraction far harder.

These infrastructure changes differ from product competition in important ways. Product choice requires active user decisions; infrastructure changes protect users who make no choice at all. Default privacy is more powerful than opt-in privacy because it protects the vast majority who never adjust settings.

The shift toward privacy-protective defaults reflects market discovery at the infrastructure level. Companies that control infrastructure (browser makers, operating system vendors, device manufacturers) have discovered that privacy features provide competitive advantage. Google implementing privacy features in Chrome, despite Google's advertising business, illustrates the competitive pressure: if Chrome does not provide privacy features, users migrate to browsers that do.

Data-Broker Opt-Out as a Patch on the Regime

A subscription market now exists for services that attempt to remove an individual's records from the hundreds of data-broker directories that aggregate public records, property filings, voter rolls, and purchase histories. Optery, Incogni, DeleteMe, Kanary, and Privacy Bee are the major vendors at the time of writing. The services automate opt-out requests to each broker, return screenshot evidence of removal, and repeat the process on a schedule because brokers repopulate records from primary sources.[22]

The existence of the market is the relevant observation. A subscription that must run for as long as the subject wants privacy from the data-broker industry is what U.S. law currently offers in place of a prohibition on aggregation itself, and the subscription model exists because the law permits the underlying aggregation. The tools work within their limits and are worth having. The Austrian point is that the market for these services is a patch on a regulatory regime, and a legal regime that did not permit the aggregation would not require the patch to exist.

The Limits of Market Response

Market responses are real but face structural obstacles. Network effects favor established platforms; users cannot easily switch when their contacts remain on surveillance platforms. The discovery process is slow; many users remain unaware of alternatives. Privacy products often remain harder to use than surveillance alternatives, limiting adoption to those who specifically prioritize privacy.

Market response also addresses only some dimensions of the surveillance problem. Companies can compete on privacy for functions where privacy-respecting alternatives exist. But market competition cannot directly remove government surveillance requirements, infrastructure-level monitoring, or the accumulation of data by entities that face no market pressure to delete it.

The claim is not that markets solve all problems instantly; the claim is narrower. Markets discover preferences through entrepreneurial experimentation, and competition tends toward serving those preferences over time. Privacy market development is early-stage. The trajectory points toward greater privacy competition, but the process is incomplete.

Market Discovery

Privacy preferences were latent until entrepreneurs created products that revealed them. Apple did not know 80 percent of users would reject tracking until ATT gave them the choice, and Signal did not know millions would adopt encrypted messaging until usability made adoption feasible. The discovery process has no predetermined endpoint, and current privacy tools are early-stage. Each generation is easier to use and more competitive with the surveillance alternatives. Markets have not solved the privacy problem; they are in the process of discovering how to solve it.

Chapter Summary

Corporate data extraction inverts the traditional market relationship. The advertiser is the customer, the user supplies raw material for prediction products, and the business competes to capture users; serving them is incidental to the model. Corporate and state surveillance have become symbiotic: legal requirements force data collection, voluntary cooperation provides government access to data companies could never have been legally required to collect directly, and the public-private partnership achieves surveillance scope neither side could accomplish alone. Current outcomes also reflect substantial state intervention: intellectual property creates

platform monopolies, compliance regulations build moats around incumbents, and government contracts incentivize surveillance development. The free-market counterfactual would produce different incentives.

Biometric and genetic data sit in a distinct category. They cannot be rotated, because a fingerprint, iris scan, face, voice print, or genome is the same identifier across every year of a person's life. A breach is permanent and irrecoverable for the individuals whose records it contained, and legal remedies can price exposure after the fact without undoing it. The defensive posture this forces is prevention-based and must succeed every time, since a single breach is permanent.

Markets are responding to the privacy demand that intervention has not satisfied. Apple's App Tracking Transparency revealed that roughly 80% of users reject tracking when given a clear choice, encrypted messaging has achieved mainstream adoption, and paid services offer alternatives to ad-supported extraction. The discovery is incomplete. Network effects, coordination problems, and government surveillance mandates create obstacles that market competition alone may not overcome, and further regulation risks entrenching existing surveillance infrastructure through compliance moats that favor incumbents. The analysis neither condemns markets nor exonerates them. It notes that current incentives are shaped by intervention as much as by consumer preference, and that technical and entrepreneurial solutions may succeed where regulatory ones would reinforce the structure they claim to restrain.[23]

Notes

[1]Neil Postman, *Technopoly: The Surrender of Culture to Technology* (New York: Alfred A. Knopf, 1992), 7. Written a decade before the surveillance-capitalism business model achieved dominance, Postman's framework anticipated the pattern this chapter analyzes: once a technology is introduced for narrow commercial purposes, its inherent logic restructures the activities around it. The advertising-supported internet did not invent user surveillance; it inherited a logic from what Postman called the technopoly, in which instruments of measurement become the frame through which experience itself is understood. For the popular-culture compression of the same argument, see Andrew Lewis

(user "blue_beetle"), comment on MetaFilter, August 26, 2010: "If you are not paying for it, you're not the customer; you're the product being sold," a formulation this chapter's analysis extends and grounds.

[2]Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, 2019), 8–11. Zuboff provides detailed documentation of surveillance capitalism's development and practices, though her analysis differs from the Austrian perspective applied here.

[3]Variable reward schedules—intermittent reinforcement patterns that maximize engagement by delivering unpredictable rewards—were first described by B.F. Skinner in *The Behavior of Organisms* (New York: Appleton-Century-Crofts, 1938). Their application to digital product design is documented in Natasha Dow Schüll, *Addiction by Design: Machine Gambling in Las Vegas* (Princeton, NJ: Princeton University Press, 2012), which analyzes how slot-machine designers use variable-ratio schedules to maximize time-on-device; the same logic transferred directly to social media feed design. Tristan Harris's work at the Center for Humane Technology, `https://www.humanetech.com/`, documents the deliberate application of these techniques to smartphone applications.

[4]On network effects and platform economics, see David S. Evans and Richard Schmalensee, *Matchmakers: The New Economics of Multisided Platforms* (Boston: Harvard Business Review Press, 2016).

[5]Richard Cantillon, *Essai sur la Nature du Commerce en Général*, ed. Henry Higgs (London: Macmillan, 1931; French original 1755), Part II, chapters 6–7. For the Austrian treatment, see Ludwig von Mises, *The Theory of Money and Credit*, trans. H. E. Batson (New Haven: Yale University Press, 1953 [1912]), Part Two, chapter 8; and Rothbard, *Man, Economy, and State* (cited at Chapter 3, note 3), chapter 11. The restriction of the Cantillon effect to monetary expansion is deliberate in the Austrian literature, and the analogy to data flows in the main text claims only that the structural pattern recurs, not that the monetary analysis transfers.

[6]The PRISM program was revealed through documents leaked by Edward Snowden in 2013. See Glenn Greenwald, *No Place to Hide: Edward Snowden, the NSA, and the U.S. Surveillance State* (cited at Chapter 1, note 7).

[7]National Security Letters are administrative subpoenas issued by the FBI under 18 U.S.C. § 2709 (electronic communications), 12 U.S.C. § 3414 (financial records), and related statutes, compelling disclosure of customer records while prohibiting recipients from acknowledging the request. The gag-order provision was challenged in *Doe v. Ashcroft*, 334 F. Supp. 2d 471 (S.D.N.Y. 2004), and subsequently in *Doe v. Mukasey*, 549 F.3d 861 (2d Cir. 2008), which required judicial review of gag orders. The FBI issued approximately 16,000–17,000 NSL requests per year in the early 2020s. For the scale of use, see Electronic Frontier Foundation, "National Security Letters," `https://www.eff.org/issues/national-security-letters`, and the annual FBI reports to Congress on NSL usage required under 18 U.S.C. § 2709(e).

[8]On the Austrian critique of market failure theory, see Israel M. Kirzner, "The Perils of Regulation: A Market-Process Approach," in *Discovery and the Capitalist*

Process (Chicago: University of Chicago Press, 1985), and Rothbard, *Power and Market* (cited at Chapter 10, note 2), chapter 3.

[9]General Data Protection Regulation (GDPR), Regulation (EU) 2016/679, Official Journal of the European Union, L 119/1 (May 4, 2016), effective May 25, 2018, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679. The GDPR imposes data-minimization, consent, and accountability obligations on firms processing EU residents' personal data. For the compliance-cost asymmetry argument—that large incumbents absorb compliance costs that exclude smaller rivals—see Cristina Caffarra and Fiona Scott Morton, "The European Commission Digital Markets Act: A Translation," VoxEU, https://cepr.org/voxeu/ (2021); and Garrett Johnson, Scott Shriver, and Samuel Goldberg, "Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR," *Marketing Science* 42, no. 3 (2023): 454–480.

[10]On Section 230 and platform immunity, see Jeff Kosseff, *The Twenty-Six Words That Created the Internet* (Ithaca, NY: Cornell University Press, 2019).

[11]Frédéric Bastiat, "Ce qu'on voit et ce qu'on ne voit pas" (1850), translated as "That Which Is Seen, and That Which Is Not Seen," in *Selected Essays on Political Economy*, trans. Seymour Cain (Irvington-on-Hudson, NY: Foundation for Economic Education, 1995). Henry Hazlitt, *Economics in One Lesson* (New York: Harper, 1946), chapter 1, states the principle directly: "The art of economics consists in looking not merely at the immediate but at the longer effects of any act or policy; it consists in tracing the consequences of that policy not merely for one group but for all groups."

[12]23andMe's October 2023 credential-stuffing breach exposed roughly seven million customer profiles; the company filed Chapter 11 in March 2025, and its genetic database was auctioned as a bankruptcy asset in 2025, confirming that genetic data is a permanent, transferable exposure that routine consumer consent does not price. 23andMe, https://www.23andme.com/. 23andMe breach disclosure (December 2023), SEC filing, https://www.sec.gov/Archives/edgar/data/1804591/000110465923124825/tm2331489d1_8k.htm. Class action settlement reporting in *Reuters* and *Bloomberg*, 2024. Chapter 11 bankruptcy filing in In re 23andMe Holding Co., Case No. 25-10625, U.S. Bankruptcy Court, Southern District of Texas (March 2025). Genetic Information Nondiscrimination Act of 2008 (GINA), 42 U.S.C. § 2000ff. On the limits of HIPAA and GINA as applied to consumer genetic databases, see Natalie Ram, "Genetic Privacy After Carpenter," *Virginia Law Review* 105 (2019).

[13]India's Aadhaar biometric identity program enrolls over 1.3 billion residents with fingerprints, iris scans, and demographic data centrally held by the Unique Identification Authority of India; architectural critiques since 2017 have identified the central database as a single point whose compromise would permanently enumerate an entire population's biometric identifiers. Aadhaar / UIDAI, https://uidai.gov.in/. On architectural critiques, see Reetika Khera et al., *Dissent on Aadhaar: Big Data Meets Big Brother* (Orient BlackSwan, 2019). On biometric cancelability, see Nalini Ratha, Jonathan Connell, and Ruud Bolle, "Enhancing Security and Privacy in Biometrics-Based Authentication Systems,"

IBM Systems Journal 40, no. 3 (2001): 614–634, and the subsequent literature on biometric template protection at the IEEE International Joint Conference on Biometrics (IJCB) proceedings. Comparable programs include Pakistan's NADRA, https://www.nadra.gov.pk/, and the EU's proposed European Digital Identity Wallet under eIDAS 2.0.

[14]Illinois BIPA (2008) is the model state biometric-privacy statute with a private right of action; several U.S. states (Texas, Washington, New York, others) have passed similar laws since 2021; legal remedies price biometric exposure after the fact but cannot prevent the exposure itself. Illinois Biometric Information Privacy Act (BIPA), 740 ILCS 14, https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=3004&ChapterID=57. Texas Capture or Use of Biometric Identifier Act (CUBI), Tex. Bus. & Com. Code § 503.001. Washington State biometric privacy law, RCW 19.375. On BIPA litigation, see *Rosenbach v. Six Flags Entertainment Corp.*, 129 N.E.3d 1197 (Ill. 2019), and *Cothron v. White Castle System, Inc.*, 216 N.E.3d 918 (Ill. 2023). Comparative analysis of state biometric-privacy laws in Future of Privacy Forum reports, https://fpf.org/.

[15]During Q3 2021, approximately 80% of iOS users opted out of tracking on major social media platforms following ATT implementation. See Statista, "App Tracking Transparency: Opt-In Rate of iOS Users 2022," and Flurry Analytics, "App Tracking Transparency Opt-In Rate: Monthly Updates."

[16]Meta CFO Dave Wehner stated the estimated $10 billion revenue impact during the Q4 2021 earnings call, February 2022. See "Facebook Says Apple iOS Privacy Change Will Result in $10 Billion Revenue Hit This Year," CNBC, February 2, 2022. Lotame estimated the total impact at $12.8 billion.

[17]DuckDuckGo is a privacy-preserving search engine that does not track users or build behavioral profiles. DuckDuckGo, https://duckduckgo.com/. Founded by Gabriel Weinberg in 2008, DuckDuckGo reported passing 100 billion lifetime queries in January 2021 and processing over 3 billion monthly queries by 2023. For traffic data, see DuckDuckGo Traffic Statistics, https://duckduckgo.com/traffic. The engine shows revealed preference for privacy-respecting alternatives even when they offer less thorough results than Google.

[18]Brave is a privacy-focused browser developed by Brave Software, Inc., founded by Brendan Eich (co-creator of JavaScript and former CEO of Mozilla). Brave, https://brave.com/. Brave blocks trackers and ads by default and offers the Basic Attention Token (BAT) system, which compensates users for opting into privacy-respecting ads. Brave reached 50 million monthly active users by 2022 and 60+ million by 2023. For usage statistics, see Brave's transparency reports at https://brave.com/transparency/. The BAT model is described in Brendan Eich and Brian Bondy, "Basic Attention Token (BAT): Blockchain Based Digital Advertising" (white paper, 2017), https://basicattentiontoken.org/static-assets/documents/BasicAttentionTokenWhitePaper-4.pdf.

[19]For the canonical Signal Foundation reference and protocol specifications, see Chapter 5, note 7. On Signal's nonprofit structure and reported growth, see Signal

Technology Foundation tax filings and Business of Apps, "Signal Revenue & Usage Statistics (2025)." The Signal Foundation operates as a 501(c)(3) nonprofit funded primarily by donations.

[20]WhatsApp adopted the Signal Protocol for end-to-end encryption in 2016, completing the rollout to all one billion users in April 2016. The integration was developed in partnership with Open Whisper Systems. See Moxie Marlinspike, "WhatsApp's Signal Protocol Integration Is Now Complete," Open Whisper Systems blog, April 5, 2016, `https://signal.org/blog/whatsapp-complete/`. The Signal Protocol specification is at `https://signal.org/docs/`. WhatsApp's adoption represents the largest deployment of end-to-end encryption in history; that a surveillance-dependent parent company (Meta) found encryption competitively necessary illustrates the market pressure this chapter analyzes.

[21]DNS-over-HTTPS (DoH) encrypts DNS queries between the client and a DNS resolver, preventing ISPs from observing domain-name lookups. The standard is defined in Paul Hoffman and Patrick McManus, "DNS Queries over HTTPS (DoH)," RFC 8484, Internet Engineering Task Force (October 2018), `https://www.rfc-editor.org/rfc/rfc8484`. Major browser deployments include Firefox (default DoH since 2020, `https://blog.mozilla.org/en/products/firefox/firefox-continues-push-to-bring-dns-over-https-by-default-for-us-users/`) and Chrome (secure DNS since 2020). ISP opposition to DoH deployment is documented in UK parliamentary evidence from Virgin Media, BT, and Sky, and in the Internet Society's analysis at `https://www.internetsociety.org/resources/doc/2019/dns-over-https/`.

[22]Optery, Incogni, DeleteMe, Kanary, and Privacy Bee are the major vendors in the data-broker opt-out subscription market; the market's existence shows that U.S. law prices data-broker aggregation through post-hoc opt-out instead of through prohibition. Optery, `https://www.optery.com/`. Incogni (Surfshark), `https://incogni.com/`. DeleteMe (Abine), `https://joindeleteme.com/`. Kanary, `https://www.kanary.com/`. Privacy Bee, `https://privacybee.com/`. For the primary regulatory document on the underlying data-broker industry, see Federal Trade Commission, *Data Brokers: A Call for Transparency and Accountability* (May 2014), already cited in note 16. For the Vermont data-broker registration regime as one state-level response, 9 V.S.A. § 2446.

[23]Further reading on surveillance capitalism, adtech, and data brokers. On the canonical critical treatment, Zuboff, *The Age of Surveillance Capitalism*, see note 2 above and Chapter 1, note 6; Carissa Véliz, *Privacy Is Power* (Bantam Press, 2020), is the shorter popular version. On the behavioral-prediction mechanism specifically, Jaron Lanier, *Ten Arguments for Deleting Your Social Media Accounts Right Now* (Henry Holt, 2018), is accessible; Yasha Levine, *Surveillance Valley: The Secret Military History of the Internet* (PublicAffairs, 2018), documents the state–commercial continuity. On the 2013 NSA revelations underlying the PRISM citation in note 5, Greenwald, *No Place to Hide* (cited at Chapter 1, note 7), is the

narrative account. On data brokers specifically, the FTC report *Data Brokers: A Call for Transparency and Accountability* (May 2014), `https://www.ftc.gov/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014`, remains the foundational regulatory document; Kashmir Hill's ongoing reporting at *The New York Times* is the best current journalism on specific data-broker practices. On the Austrian critique of "market failure" applied to platforms, Kirzner, "The Perils of Regulation" (cited in note 6 above), is the theoretical foundation; for empirical work, Luigi Zingales, *A Capitalism for the People: Recapturing the Lost Genius of American Prosperity* (Basic Books, 2012), and the ongoing Stigler Center working papers at `https://stigler.capitalisnt.com`, treat rent-extraction in platform markets. On the tracking-mechanics side, Timothy Libert, "Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites," *International Journal of Communication* 9 (2015): 3544–3561, is the standard empirical study; Arvind Narayanan and colleagues' OpenWPM work at Princeton, `https://webtransparency.cs.princeton.edu`, is the technical-measurement reference. For the hedonic-tradeoff critique of "free services," Hal Varian, "Beyond Big Data," *Business Economics* 49, no. 1 (2014): 27–31, is the industry-defender view worth reading alongside the critics.

12

THE ANALYTICS STACK

"A government bureau's interest in data does not
end with what it is lawfully authorized to collect."
Anonymous, origin unknown

Commercial firms with no legislative mandate have built surveillance capability at a scale no individual agency could justify on its own, and state agencies purchase the output. Sensors harvest, brokers aggregate, models analyze, and agencies deploy. What the Constitution forbids any single actor to do, the market assembles in stages for the state to buy. Surveillance has migrated from direct observation by the state into an industrial process with priced layers at every step. Chapter 13 shows the state attacking cryptographic tools; this chapter shows the state buying the observational layer that exists before content is encrypted and after it is decrypted.

12.1 The Division of Labor in Surveillance

Adam Smith observed that the division of labor increases productivity by allowing specialization.[1] Ludwig von Mises extended this to show that specialization depends on market exchange: the division of labor is coextensive with the catallactic order.[2] The Analytics Stack applies that principle to observation.

A single agency cannot legally build and maintain a facial-recognition database containing every photograph on the open web. No single agency has the mandate or the political cover for that. A private firm can build the same database under the far weaker

constraints that govern commercial speech, and then sell access to the agencies that could not have built it directly. The agency pays a subscription. The firm pays for the engineering. Neither bears the full reputational or constitutional cost a direct program would have incurred.

The same logic applies across every layer of the stack. License-plate capture, location tracking, social-graph extraction, behavioral prediction, endpoint compromise, and analytical fusion are each performed by specialist firms pricing their output in a competitive market. The state is one customer among many. That posture allows the firm to claim that it is neither a government contractor in the regulatory sense nor a state actor for constitutional purposes, and it allows the state to pay market price for capabilities it could not legally assemble.[3]

The arrangement is the predictable consequence of two doctrines interacting. The first doctrine is the third-party doctrine, which holds that information voluntarily shared with a third party enjoys reduced Fourth Amendment protection. The second doctrine is the commercial-speech permission for data collection about individuals who are not customers of the collector, which covers web scraping, ad-tech bidstream harvesting, and license-plate photography on public roads. Between them lies a market for the state's observational appetite, with the first doctrine authorizing acquisition and the second doctrine manufacturing the supply. Better rules will not patch this; the rules produced it.[4]

The Coalition the Stack Assembles

Every state activity divides the population into a class that pays more into the apparatus than it receives and a class that receives more than it pays. Rothbard's innovation was to include in the second class not only state employees but the contractors, vendors, and recipients of monopoly grants whose revenues depend on the state's operation continuing.[5] The firms that compose the analytics stack belong to that class. A data broker selling to the Department of Homeland Security, a facial-recognition vendor selling to Immigration and Customs Enforcement, a spyware firm selling to allied

ministries of interior, and a predictive-policing platform selling to municipal departments are each revenue-aligned with the state's continued purchase.

The stack's durability under political pressure follows from this alignment. The firms' incomes depend on the state's continued purchase, and the state's continued purchase depends on the existence of the capability the firms produce. Public pressure against any single contract reaches a coalition whose members absorb the pressure together, and the coalition is durable because each member has commercial incentive to keep the others solvent.

12.2 Machine Vision at Scale

Facial recognition is the mechanism by which a visual identifier that was once bounded by human memory becomes an index into a permanent database. The mathematical transformation is ordinary: a neural network maps an image to a vector in a high-dimensional space, and two images of the same face produce vectors that are close under a chosen metric. The transformation was well-understood long before it became a political question. What changed was cost.

Image acquisition fell to near-zero through ubiquitous camera deployment. Model inference fell to fractions of a cent through consumer GPU economics. Storage of feature vectors fell to trivial cost through commodity databases. Once all three costs had collapsed, matching a photograph against a billion-image corpus became cheaper than the paperwork of not doing so, and a capability that had been theoretically available became operationally universal.

Clearview AI is the canonical case because it removed the one remaining constraint on the capability: corpus size. By scraping images from the open web, Clearview built a corpus of roughly twenty billion photographs. That corpus is larger than any government-assembled database and subject to no legislative oversight. Law-enforcement agencies that could not have lawfully assembled such a corpus subscribe to it as a service. European data-protection authorities in several member states have imposed fines and bans, with no operational effect inside the United States. The ACLU

settlement in the Illinois Biometric Information Privacy Act case restricts Clearview's U.S. commercial sales to government entities, which is the clientele the firm was already serving.

The principle does not depend on Clearview specifically. Any firm willing to ignore scraping norms can build a comparable corpus, and the replaceability of the firm is what makes the capability durable. The Department of Homeland Security's Homeland Advanced Recognition Technology (HART) database is designed to hold biometric and demographic records for hundreds of millions of identities, integrated with the Customs and Border Protection biometric entry-exit system and the Transportation Security Administration's airport facial-matching kiosks that are now deployed at dozens of major U.S. airports. A single identity check that was never legislated as such now runs across commercial corpora, federal indices, airport infrastructure, and border crossings.[6]

Human recognition had a natural cost ceiling, because memory is finite and a passerby forgets within minutes. Machine recognition has no such ceiling. A capability whose use was rationed by cost is no longer rationed once the cost collapses, and it gets deployed at whatever rate other constraints permit. The law written for the scarce regime does not describe the abundant one.

12.3 The Model as Analyst

The labor-intensive step in historical intelligence work was reading. Documents accumulated faster than analysts could process them, and most collection was dead weight for that reason. Large language models change the accounting. A single analyst with a suitably prepared model can now produce summaries, extractions, translations, and correlations across document volumes that previously required organizational-scale human attention. The bottleneck has moved from reading to defining the query.

Palantir's Artificial Intelligence Platform, launched in 2023 and bundled with the firm's existing Gotham and Foundry products, is the visible instance of this shift. By 2025 its revenue exceeded four billion dollars, with contracts across the Department of Defense,

the Department of Homeland Security, Immigration and Customs Enforcement, and the Navy. Contract values matter less than what they purchase. What they purchase is a fused data environment with a model-driven query interface, in which an analyst asks a natural-language question and receives an answer drawn from structured records, unstructured documents, satellite imagery, and sensor feeds. A multi-day collaboration among specialists collapses into a single conversational exchange.[7]

The book's earlier argument was that privacy defense is cheaper than privacy offense, because encryption costs cents while breaking it costs billions. That argument holds for each unit of communication. The analytics stack does not attack individual units; it attacks aggregation. A million intercepted records are worth nothing if no analyst can read them. A million intercepted records fed to a model are worth everything, because the model extracts the patterns at a cost per query that approaches zero. The cost asymmetry that once favored defense has been partly offset at the aggregation layer.

The target of defense has moved. Encrypting content remains essential, but it is no longer sufficient. What must also be protected is the metadata that the model uses to build patterns when content is unavailable: timing, social graph, location, device identifier, and frequency. The model is a metadata analyst first and a content analyst only when metadata fails, which makes the transport-level defenses Chapter 17 develops indispensable.

12.4 The Data-Broker Channel

A data broker is a firm that aggregates records about individuals and resells them. The industry predates the internet but scaled with it, and its economic function was settled before anyone thought to ask whether it should exist. Brokers buy from app publishers, loyalty programs, credit bureaus, public records, and the real-time bidding stream through which advertising auctions run. They sell to advertisers, insurers, employers, private investigators, and, through specialist intermediaries, to government agencies.

The government-sales channel is the feature of interest. The Supreme Court's 2018 decision in *Carpenter v. United States* held that the government needs a warrant to compel a carrier to produce historical cell-site location information. The decision was narrowly reasoned. The court emphasized the carrier's compelled retention and the detailed nature of historical cell-site records. No part of the decision addressed information the government does not compel but purchases from a commercial party that lawfully obtained it. That gap is the data-broker loophole.

Venntel, a subsidiary of the advertising-tech firm Gravy Analytics, sold location data drawn from the real-time bidding stream to federal agencies across the late 2010s. The Treasury Inspector General for Tax Administration confirmed in a 2023 report that Internal Revenue Service Criminal Investigation had used Venntel data without warrants between 2017 and 2018. These were domestic tax-enforcement matters, not national-security exceptions. In 2024 the Venntel capability was absorbed by Babel Street and rebranded as Locate X. Fog Reveal, another product in the same class, was documented by the Electronic Frontier Foundation and the Associated Press as being used by dozens of local, state, and federal agencies at every level of U.S. law enforcement.[8]

A Fourth Amendment written against direct state acquisition, whether by search or by seizure or by compulsion, has no language for the case in which the state does none of these and only buys. The information the state obtains through purchase is as detailed as anything a warrant could have produced, and the legal theory under which the state obtained it is that no one required a warrant to buy what the market already sold. Congress has repeatedly considered bills to close the loophole through a Fourth Amendment Is Not for Sale Act, each of which has failed in committee or died on the calendar, leaving the commercial channel in continuous operation.

Chapter 10 analyzed financial surveillance as triangular intervention in Rothbard's sense: the state conditions banking licensure on the bank reporting on its depositors, and the terms of the bank-depositor exchange are coercively modified to include state observation. The data-broker channel reaches the same surveillance

outcome without triangular intervention. Every transaction in the chain is voluntary. The subject exchanges with an app publisher, the publisher resells to a broker, the broker sells to the state. The state compels none of these exchanges and only purchases their outputs. What does the work is not coercion of an exchange but selective absence of protection around it. The third-party doctrine withdraws the subject's Fourth Amendment interest once data leaves the subject's hands, and the commercial-speech framework permits the resale the doctrine has made legally riskless. Same information flow as the compelled case; different mechanism; no warrant required. The state that once needed to compel banks has discovered that a legal regime producing a private surveillance market is the cheaper instrument.

12.5 Physical-World Infrastructure

The stack is not confined to the internet. The same cost curves that made facial recognition economic have made every form of physical-world sensing economic.

Automated license-plate recognition (ALPR) is the paradigmatic case because a license plate is the one identifier every automobile is required by law to carry, and its retention in law is justified as a minimal mechanism for traffic enforcement. Flock Safety, the leading vendor, operates ALPR infrastructure in over five thousand communities across nearly every state, capturing billions of vehicle scans per month. Each scan is an image, a plate read, a location, and a timestamp. Aggregated over months, the scans produce a movement log for any vehicle that traverses the covered jurisdictions. The Norfolk, Virginia Circuit Court ruled in 2024 that warrantless access to this movement log violates the Fourth Amendment, and the ruling will be litigated upward for years; during that time, the infrastructure continues to scan.

The same vendor acquired a police-drone manufacturer in late 2024 and is building a Drone-as-First-Responder product line. The DFR pattern, pioneered by Chula Vista Police in California and now deployed by hundreds of U.S. cities, dispatches a drone to

any reported incident before the officer arrives. The drone records continuous video during its flight, and the recording enters the evidentiary system whether the incident produced a crime or not. Skydio and BRINC build U.S. platforms for this market. DJI's Chinese origin has pushed it out of federal procurement. The capability is agnostic to vendor.

Ring, the Amazon-owned home-camera product, illustrates the partial-retreat variant. From 2018 through 2024 the Neighbors app allowed police to solicit footage directly from Ring owners through a "Request for Assistance" tool. Public opposition to the surveillance implications led Amazon to remove the tool in January 2024. The infrastructure did not disappear. Police now obtain Ring footage through warrants served on Amazon or through owner consent, which is the older process with a higher transaction cost. Public attention can raise the price of a surveillance channel without closing it.

Cell-site simulators, historically branded as Stingrays and now sold under various successor names, complete the physical layer. A cell-site simulator masquerades as a cellular tower. Phones in the area connect to it by default and disclose their International Mobile Subscriber Identity. The simulator can log presence and track movement, and in more intrusive configurations it can intercept metadata directly. Research published in 2023 and 2024 documented that cell-site simulators remain operationally viable against 5G non-standalone networks and against some 5G standalone configurations. Mitigations in the 3GPP standardization track exist but lag deployment by years; in the interim the capability persists.[9]

The physical-world infrastructure inverts one of the oldest implicit privacy assumptions. The common-law expectation that a person was observable in public was built around the cost ceiling of human observation. One officer could follow one subject. A city could not follow everyone. Cameras and readers and drones removed the cost ceiling. The expectation that survived in common law has no economic basis in the new regime. Privacy in public is now an engineering problem.

12.6 The Commercial Spyware Arm

The analytics stack assumes the target's device is a reliable source of the target's data. That assumption fails when the device is compromised. Commercial spyware has matured into an industry with vendor tiers, standard product catalogs, published deployment records, and an active accountability case law.

Pegasus, from the NSO Group, is the reference product. It deploys through zero-click exploits against mainstream messaging endpoints including iMessage and WhatsApp, leaving the target unaware that the compromise has occurred. Citizen Lab at the University of Toronto has documented Pegasus infections of journalists, dissidents, lawyers, and politicians across at least thirty countries, including the first confirmed infection of a Russian journalist on European Union soil in 2023. The United States Commerce Department placed NSO on the Entity List in November 2021, which restricted U.S. technology sales to the firm without shutting down its operations. In late 2024 a U.S. federal court ruled NSO liable under the Computer Fraud and Abuse Act for the infection of roughly 1,400 WhatsApp users. That was the first major civil accountability decision in the spyware market.

Intellexa, the consortium behind the Predator spyware product, was sanctioned by the U.S. Treasury in March 2024 after its founders were identified and its corporate shell traced. Paragon Solutions, which markets the Graphite product, was acquired by the U.S. private-equity firm AE Industrial Partners in December 2024, producing a domestic-owned spyware vendor. Candiru and several smaller firms occupy adjacent market segments. Products differ in exploit portfolio and target-platform coverage and price; the capability is fungible, and the industry's survival of sanctions on individual firms shows that restricting vendors does not restrict the capability. The competitive structure reproduces the Clearview pattern: sanction one vendor, another takes the market.

Forensic-extraction tools occupy the parallel wide-deployment layer. Cellebrite and Magnet Forensics (which acquired Grayshift and its GrayKey product in 2023) sell devices that exploit boot-chain and lockscreen weaknesses to extract data from seized phones. The

customer base is law-enforcement agencies worldwide, numbering in the thousands. Unlike Pegasus, these tools require physical possession of the device, but they are priced for routine police use, and they operate at the scale of ordinary criminal investigations. The marginal arrest now generates a full device extract as a matter of course.

GrapheneOS is the public exception. A hardened Android distribution running on Pixel hardware, it auto-reboots into a Before First Unlock state after a configurable idle period, restricts USB peripherals while locked, and supports a duress wipe. Leaked Cellebrite support matrices from 2024 listed GrapheneOS devices as difficult or inaccessible in configurations that defeat stock Android and iOS. Extraction remains possible under a sufficiently resourced attack, but the floor the user can set is high enough that routine extraction fails.

Executive Order 14093, signed by President Biden in March 2023 and continued under successor administrations, prohibits U.S. government operational use of commercial spyware that poses counterintelligence or human-rights risks. The order is the state's own acknowledgment that the capability is dangerous when directed at its officials. No restriction accompanies the capability when directed at anyone else. State actors want the spyware when the targets are adversaries and want to restrict it when the targets are themselves.[10]

Chapter 14 will set this limit out in full: cryptography cannot solve endpoint compromise, because a compromised device reads cleartext the same way the user does. The analytics stack sits on that limit. Breaking modern encryption costs billions in compute or the discovery of a zero-day in the implementation; compromising an endpoint costs tens of thousands to low millions per target for remote spyware and tens of thousands per unit for physical extraction, and delivers cleartext directly. The arms race migrated to the layer where the price structure favored the attacker, and it has stayed there. Chapter 22 takes up the countermeasures.

12.7 Compressing the OODA Loop

The first chapter of this book introduced Boyd's observe-orient-decide-act loop as the structure of any targeted intervention, and argued that privacy disrupts the loop at observation. The Analytics Stack is useful to examine as the adversary's answer to that claim, because the stack compresses the entire loop into industrial process.[11]

The observation layer is the sensor and acquisition stack: ALPRs, doorbell cameras, cell-site simulators, scraped web images, app location streams, bank transaction reports, and commercial spyware. Its output is raw records.

The orientation layer is the fusion and analysis stack: Palantir-class platforms, broker-assembled dossiers, biometric matching services, and large language models querying the fused environment. Its output is hypotheses and patterns, organized by subject or by event.

The decision layer is the case-management and authorization stack: warrant requests, arrest authorizations, audit selections, travel restrictions, and administrative sanctions. Its output is enforcement orders directed at individuals, and it has automated alongside the others. Pretrial risk scoring through the Public Safety Assessment and COMPAS shapes detention decisions in hundreds of U.S. jurisdictions. IRS audit selection runs on statistical models that choose which returns receive human review. ICE's Risk Classification Assessment was documented in 2018 to recommend detention in every case evaluated, the failure mode of any decision-layer automation not empirically checked. Australia's Robodebt issued automated welfare-debt notices against hundreds of thousands of recipients before courts halted the underlying methodology in 2019, and a 2023 Royal Commission judged the scheme unlawful and harmful. The pattern is consistent: models recommend, humans sign, and the review window shrinks to seconds per case as throughput rises.[12]

The action layer is deployment of enforcement: police units, tax audits, asset freezes, border stops, and regulatory actions. Its output is intervention in the target's life.

Every layer in that sequence has specialist vendors and commercial pricing and has been optimized for throughput. The compressive effect is that an officer who in 1990 would have needed weeks of investigative work to assemble a case now has a query interface that returns the same package in minutes. The compression affects both sophisticated and routine operations. Cases previously uneconomic to pursue are now worth pursuing, which expands the population brought into contact with enforcement.

The countermeasure logic follows directly. If the stack compresses the loop, defense must act earlier in the loop. Observation-layer defenses are the only ones that break the entire chain, because a fused platform cannot fuse what was never collected. Encryption denies the observation layer the content; location obfuscation denies the observation layer the movement log; pseudonymity denies the observation layer the identity; cash denies the observation layer the transaction. Each of these defenses becomes more important as the downstream layers improve, because the downstream layers are the multiplier that turns observation into action. The defender's marginal sensor denial is worth what it is worth times the multiplier of every downstream improvement.

Risk Is Not Uncertainty

Frank Knight's 1921 distinction illuminates what the compressed loop can and cannot deliver.[13] Risk is the condition under which the probabilities of outcomes are known or empirically estimable. Under uncertainty, those probabilities are not known. Insurance handles risk; no formula handles uncertainty. Knight's distinction, developed outside the Austrian tradition and adopted by later Austrians including Lachmann and Shackle, is the right vocabulary for naming the limit.

A pattern recurring in historical data is a risk: what happened before can happen again at roughly the same rate, and a fused analytical platform can estimate the rate and direct resources at the higher-rate cases. A novel event of a kind the data does not contain is uncertainty: no aggregation reduces the novelty, because the novelty is in the category itself and not in any parameter within

a known category. The compressed OODA loop improves enforcement against the pattern class and does not improve enforcement against the novel class. State threat models that depend on the novel class receive no benefit from the aggregation despite its cost. The marketing conflates the two classes because the political constituency for the stack does not distinguish them, and the epistemics do. Surveillance infrastructure priced against Knightian uncertainty is priced against a problem it cannot touch.

12.8 Why Agencies Keep Buying

Mises analyzed the dynamics of bureaucratic management in his 1944 study of the phenomenon.[14] Bureaucratic management is the management of organizations whose output cannot be evaluated through profit and loss, because the output has no market price. An agency that loses money on its operations does not, by that fact alone, change course, and an agency that produces more intelligence reports per unit of budget does not thereby become better at its function in any measurable sense. Absent the profit test, what expands is the budget and what defines success is the budget's expansion.

Applied to the analytics stack, the argument names the internal dynamic that the cost-curve analysis alone does not. Cost curves explain why the stack became affordable. Mises's analysis of bureaucracy explains why affordable capability is purchased and retained regardless of whether it produces useful output. A director who declines to buy a commercial sensor feed that peer agencies have bought faces a political cost if any incident occurs that the feed could have prevented. A director who buys the feed faces no corresponding cost for the feed producing no usable intelligence, because no metric inside the agency measures the marginal return on the purchase. The asymmetry favors purchase and favors retention, and it holds regardless of what the sensor accomplishes.

Bureaucracies also expand along the axes the budget permits. An agency with surplus budget acquires capability; an agency with new capability acquires staff to operate it; staff acquire promotions

and projects that require the capability to remain in place. The system locks in at the new level, and the next budget cycle begins from a baseline that includes the new level. The stack's economics gave every agency with budget authority a reason to acquire. The Misesian analysis explains why the acquisitions accumulate instead of being pruned.

12.9 Why the Stack Exists Now

The stack is not new in kind. Pinkerton's agency assembled surveillance for industrial and state clients in the nineteenth century.[15] The Stasi assembled files on a third of East German adults with 1970s technology.[16] What is new is the price per subject and the total addressable population, and both changed for identifiable reasons.

Sensor cost broke first. Digital imaging replaced film, commodity silicon replaced special-purpose hardware, cloud storage replaced local storage, and ubiquitous wireless replaced wired-only deployment. The cost of capturing a given observation fell by roughly six orders of magnitude over forty years. Storage cost broke next. Data-warehousing and object storage combined with Moore's scaling law in density to reduce the cost of retaining a given record indefinitely from economically prohibitive to economically irrelevant. Analysis cost broke last. Human analysts gave way to rule-based systems, and rule-based systems gave way to large language models, with the cost of drawing an inference from a given record falling by roughly three orders of magnitude between 2010 and 2025.

When more than one cost curve collapses simultaneously, a capability that had been available only for exceptional cases becomes available for routine ones. The stack assembled itself through ordinary competitive pressure: any agency that refused it was outperformed by agencies that adopted it, and the outperformed agencies either adopted or lost budget and jurisdiction to the ones that had. The equilibrium emerged from local choices under iden-

tical incentive gradients, which is the signature of market order operating against the user.

The economic implication for privacy defense is that the defender's cost curve must also bend. This is possible, and it has been happening, but it has been happening at a slower rate than the attacker's stack for most of the last decade. Tor, Signal, Bitcoin, and Nostr are the points on the defender's curve that have bent fastest.[17] Each is a structural innovation that reduces defender cost by orders of magnitude for a specific observation type. The pace of those innovations is the variable that determines whether the Analytics Stack will be the end of the story or an intermediate phase before defense catches up.

Chapter Summary

The Analytics Stack is the integrated system through which commercial firms assemble surveillance capability and state agencies deploy it. Four layers compose it, sensor acquisition, fusion and analysis, decision and authorization, and enforcement action, and every layer has specialist vendors pricing their output competitively. Surveillance has been industrialized through market specialization: capability no single agency could lawfully build is assembled by firms and purchased as a service, and the state pays market price for capabilities it could not have obtained directly. The commercial channel routes around constitutional protections designed against direct state acquisition. The Fourth Amendment was written against compelled records; *Carpenter* narrowly limited warrantless cell-site data without addressing the broader commercial market in equivalent information, and the gap is where the data-broker industry operates.

Cost curves explain the timing. Sensor cost, storage cost, and analysis cost all collapsed in overlapping phases, and when they broke simultaneously the stack became economically mandatory for any agency whose competitors had adopted it. The result compresses Boyd's OODA loop into industrial process: every stage from raw record to enforcement has been made efficient at once.

Defense must therefore act at the observation layer, because every downstream layer has been optimized for throughput. Encryption, anonymous routing, location obfuscation, pseudonymity, and cash each deny the stack one specific observation class, and the defender's marginal gain from any one of them is multiplied by the downstream efficiency.

The stack is not invincible. Defender cost curves can bend downward too, and the structural innovations the rest of the book develops, strong encryption, anonymous routing, resistant money, decentralized social infrastructure, have track records of collapsing the attacker's cost advantage for specific observation types. The stack's components also differ in reach and cost: a license-plate reader produces a movement log while commercial spyware produces total device compromise, and Chapter 22 threat-models against each layer separately. Legal reform has a place but not the primary one, because regulatory pressure on any single component displaces the capability to an adjacent one. Individual action is not futile: the stack is economically mandatory at the aggregate level and structurally fragile at the individual level, and a target who denies observation at every layer is a target the downstream machinery cannot process. The race has moved from the single-communication layer up to the aggregation layer, and the rest of the book works at the altitude the adversary now occupies.

Notes

[1]Adam Smith, *An Inquiry into the Nature and Causes of the Wealth of Nations* (London: W. Strahan and T. Cadell, 1776), Book I, chapters 1–3, on the division of labor, its productivity effects, and its dependence on the extent of the market. The canonical modern edition is Adam Smith, *The Wealth of Nations*, ed. R. H. Campbell, A. S. Skinner, and W. B. Todd, Glasgow Edition of the Works and Correspondence of Adam Smith, vols. 1–2 (Oxford: Oxford University Press, 1976; reprinted Indianapolis: Liberty Fund, 1981).

[2]Ludwig von Mises, *Human Action: A Treatise on Economics*, 4th ed. (Irvington-on-Hudson, NY: Foundation for Economic Education, 1996; originally published 1949), chapter 8 ("Human Society"), especially section 3 ("The Division of Labor") and section 5 ("The Catallactic Problem of the Division of Labor"). Mises argues that the division of labor is constituted by and coextensive with the market exchange order (what he terms the catallactic order), and that the

productivity gains Smith identified are themselves a product of voluntary cooperation under the price system. The surveillance-stack argument inherits this framework: specialization in observation is possible and economically rational because market prices coordinate each layer.

[3]Byron Tau's 2024 book is the standing journalistic map of the commercial-to-state surveillance market; the EFF Atlas of Surveillance and 404 Media carry the ongoing reporting. Byron Tau, *Means of Control: How the Hidden Alliance of Tech and Government is Creating a New American Surveillance State* (Crown, 2024). Electronic Frontier Foundation, *Atlas of Surveillance*, `https://atlasofsurveillance.org/`, tracks specific deployments by jurisdiction. 404 Media, `https://www.404media.co/`, publishes ongoing investigative reporting on the industry.

[4]Carpenter requires a warrant for compelled carrier records but does not address commercial purchase; legislative attempts to close the loophole through a "Fourth Amendment Is Not for Sale Act" have not passed. *Carpenter v. United States*, 585 U.S. 296 (2018). Scholarly treatment in Orin S. Kerr, "The Case for the Third-Party Doctrine," *Michigan Law Review* 107 (2009): 561–601, and Paul Ohm, "The Fourth Amendment in a World Without Privacy," *Mississippi Law Journal* 81, no. 5 (2012): 1309–1355. Legislative response in the "Fourth Amendment Is Not for Sale Act," tracked across multiple Congresses at `https://www.congress.gov/`.

[5]Murray N. Rothbard, *Anatomy of the State* (Auburn, AL: Ludwig von Mises Institute, 2009; first published 1974 in *Egalitarianism as a Revolt Against Nature and Other Essays*), in which Rothbard revives John C. Calhoun's distinction between net taxpayers and net tax-consumers. The modern restatement: every state activity produces a class that receives more from the state than it contributes and a class that contributes more than it receives, and the ruling class in Rothbard's sense is the coalition of the former. See also *Power and Market: Government and the Economy*, 4th ed. (Auburn, AL: Ludwig von Mises Institute, 2006), for the same analysis applied to specific intervention types; and John C. Calhoun, *A Disquisition on Government* (1851), for the original distinction Rothbard extended. The class analysis here is structurally distinct from the Marxist version because the class boundary is defined by state privilege, not by ownership of means of production.

[6]Clearview AI built the canonical open-web facial-recognition corpus (~twenty billion images); HART is the parallel federal biometric index; TSA and CBP deploy facial matching at the physical checkpoints. Clearview AI, `https://www.clearview.ai/`. Kashmir Hill, *Your Face Belongs to Us: A Secretive Startup's Quest to End Privacy as We Know It* (Random House, 2023), is the definitive journalistic account. U.S. Department of Homeland Security, Office of Biometric Identity Management (HART), `https://www.dhs.gov/obim`. TSA facial-matching at `https://www.tsa.gov/digital-id`. CBP biometric entry-exit at `https://www.cbp.gov/travel/biometrics`.

[7]Palantir's Artificial Intelligence Platform is the commercially visible instance of LLM-assisted intelligence analysis at state scale; Gotham and Foundry are the fused data environments it runs on top of. Palantir Artificial Intelligence Platform,

https://www.palantir.com/platforms/aip/; Gotham, https://www.palantir.com/platforms/gotham/; Foundry, https://www.palantir.com/platforms/foundry/. Federal contract values are searchable at https://www.usaspending.gov/. Critical reporting by Caroline Haskins at *Business Insider* and by Sam Biddle at *The Intercept* (https://theintercept.com/).

[8]Venntel and its successors (Locate X under Babel Street) and Fog Data Science sell location data drawn from the real-time bidding stream to federal, state, and local agencies; a 2023 TIGTA audit confirmed IRS Criminal Investigation used Venntel data without warrants. Venntel (subsidiary of Gravy Analytics), https://venntel.com/. Babel Street / Locate X, https://www.babelstreet.com/. Fog Data Science / Fog Reveal, https://fogdatascience.com/. Treasury Inspector General for Tax Administration, "The Criminal Investigation Division's Use of Commercial Databases Containing Personal Information Complied with Policy but Raised Privacy Concerns," Report No. 2023-IE-R003 (2023), https://www.tigta.gov/. Electronic Frontier Foundation reporting on the broker channel and Fog Reveal at https://www.eff.org/issues/street-level-surveillance/fog-data-science.

[9]Flock Safety dominates ALPR; Skydio, BRINC, and DJI supply the police-drone market; Amazon Ring carries the doorbell-camera layer; cell-site simulators close out the physical sensor tier. Flock Safety, https://www.flocksafety.com/. Skydio, https://www.skydio.com/. BRINC, https://www.brinc.com/. DJI, https://www.dji.com/. Amazon Ring, https://ring.com/, and the Neighbors app, https://shop.ring.com/pages/neighbors-app. Chula Vista Police Drone-as-First-Responder program, https://www.chulavistapd.org/police-services/drones-as-1st-responders. *Crowder v. Diamondback Investment Group, LLC*, Circuit Court of the City of Norfolk, Case No. CL24-000194 (2024), on warrantless ALPR access. On cell-site simulators and 5G, see Citizen Lab's ongoing analysis at https://citizenlab.ca/.

[10]Pegasus (NSO), Predator (Intellexa), Graphite (Paragon), and Candiru's products make up the remote-compromise market; Cellebrite and Magnet Forensics (GrayKey) run the physical-extraction market; GrapheneOS is the public defensive exception; Executive Order 14093 restricts U.S. government operational use without restricting the industry. Citizen Lab, https://citizenlab.ca/, is the primary empirical source on spyware deployments. NSO Group, https://www.nsogroup.com/. *WhatsApp LLC v. NSO Group*, Case No. 4:19-cv-07123 (N.D. Cal. 2024). Intellexa / Predator: U.S. Treasury OFAC sanctions designation of March 5, 2024, https://ofac.treasury.gov/recent-actions/20240305. Paragon Solutions / Graphite (acquired by AE Industrial Partners, December 2024); public reporting in *TechCrunch* and *Haaretz*. Candiru: no public website; documented by Citizen Lab and Microsoft Threat Intelligence Center reporting. Cellebrite, https://cellebrite.com/. Magnet Forensics,

https://www.magnetforensics.com/. GrapheneOS, https://grapheneos.org/, feature documentation at https://grapheneos.org/features. Executive Order 14093, "Prohibition on Use by the United States Government of Commercial Spyware That Poses Risks to National Security," 88 Fed. Reg. 18957 (March 27, 2023), https://www.federalregister.gov/documents/2023/03/30/2023-06730/prohibition-on-use-by-the-united-states-government-of-commercial-spyware-that-poses-risks-to.

[11]Boyd's OODA loop is the structural framework the chapter uses to show how surveillance integrates across layers; see Chapter 1, note 4, for the primary citation to Boyd's "Patterns of Conflict" briefing. Frans P. B. Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd* (Routledge, 2006), is the standard academic treatment. Grant T. Hammond, *The Mind of War: John Boyd and American Security* (Smithsonian, 2001), is the biographical treatment.

[12]The decision layer has automated alongside sensor and orientation layers; PSA and COMPAS drive detention decisions, ICE's Risk Classification Assessment recommended detention in nearly every case, and Australia's Robodebt generated hundreds of thousands of unlawful automated debt notices before a 2023 Royal Commission halted it. Public Safety Assessment (Arnold Ventures), https://advancingpretrial.org/psa/about-the-psa/. COMPAS / Equivant, https://www.equivant.com/. Critical analysis of COMPAS risk scoring: Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias," *ProPublica* (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. On IRS automated audit selection, see Internal Revenue Service, *Inflation Reduction Act Strategic Operating Plan* (2023), https://www.irs.gov/pub/irs-pdf/p3744.pdf. ICE Enforcement and Removal Operations Risk Classification Assessment: Reade Levinson, Mica Rosenberg, and Kristina Cooke, "Special Report: How Trump Administration Altered ICE's Detention Tool," *Reuters* (June 26, 2018), https://www.reuters.com/investigates/special-report/usa-immigration-detention/. Robodebt: Commonwealth of Australia, *Report of the Royal Commission into the Robodebt Scheme* (July 2023), https://web.archive.org/web/20230901000000/https://robodebt.royalcommission.gov.au/publications/report.

[13]Frank H. Knight, *Risk, Uncertainty, and Profit* (Boston: Houghton Mifflin, 1921), especially Part III. Knight is a Chicago School founder, not an Austrian, and the distinction the book's argument invokes is Knight's, not Mises's. Mises offers the parallel distinction between class probability and case probability in *Human Action* (cited at Chapter 3, note 1), chapter 6 ("Uncertainty"). For the Austrian reception of Knightian uncertainty, see Ludwig M. Lachmann, *Capital, Expectations, and the Market Process* (Kansas City: Sheed Andrews and McMeel, 1977), and G. L. S. Shackle, *Epistemics and Economics* (Cambridge University Press, 1972).

[14]Ludwig von Mises, *Bureaucracy* (New Haven: Yale University Press, 1944; reprinted Indianapolis: Liberty Fund, 2007), especially the chapters on "Profit

Management" and "Bureaucratic Management." Mises's argument is that bureaucracies exist where market prices cannot be formed for the outputs in question, that bureaucratic personnel are therefore not subject to the discipline profit-and-loss accounting imposes, and that the institutional form has internal expansionary pressures that no amount of administrative reform neutralizes. William A. Niskanen, *Bureaucracy and Representative Government* (Chicago: Aldine-Atherton, 1971), develops the complementary public-choice analysis from a non-Austrian vantage.

[15]The Pinkerton National Detective Agency, founded by Allan Pinkerton in 1850, operated as a private intelligence and enforcement service for railroads, industrial employers, and federal agencies throughout the nineteenth century, including strike-breaking and infiltration of labor organizations. Its client base shows that the commercial-to-state surveillance market predates digital technology by more than a century. Frank Morn, *The Eye That Never Sleeps: A History of the Pinkerton National Detective Agency* (Bloomington: Indiana University Press, 1982), is the standard history. See also Robert P. Weiss, "Private Detective Agencies and Labour Discipline in the United States, 1855–1946," *Historical Journal* 29, no. 1 (1986): 87–107.

[16]The East German Ministry for State Security (Stasi) maintained files on an estimated 5.6 million individuals (roughly a third of the adult GDR population) using approximately 91,000 full-time employees and 170,000–190,000 informal collaborators at its peak in the 1980s. The comparison with the digital analytics stack inverts the cost structure: the Stasi achieved high penetration only through massive human-labor expenditure, while the commercial stack achieves comparable or greater coverage at near-zero marginal cost per subject. Gary Bruce, *The Firm: The Inside Story of the Stasi* (Oxford: Oxford University Press, 2010). Jens Gieseke, *The History of the Stasi: East Germany's Secret Police, 1945–1990*, trans. David Burnett (New York: Berghahn Books, 2014). The Stasi Records Agency (BStU) archives are held by the Federal Commissioner for the Records of the State Security Service of the Former German Democratic Republic, `https://www.bstu.de/`.

[17]Tor provides low-latency anonymous routing through onion encryption across a volunteer relay network; Signal provides end-to-end encrypted messaging under the Signal Protocol; Bitcoin provides a pseudonymous payment system resistant to financial surveillance at the transaction layer; Nostr provides a censorship-resistant social-graph and messaging protocol over public-key cryptography. Each collapses the attacker's cost advantage for a specific observation class. Tor Project, `https://www.torproject.org/`; Roger Dingledine, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium* (2004), `https://svn.torproject.org/svn/projects/design-paper/tor-design.pdf`. Signal Protocol specification: Trevor Perrin and Moxie Marlinspike, "The Double Ratchet Algorithm" (2016), `https://signal.org/docs/specifications/doubleratchet/`; Signal Foundation, `https://signal.org/`. Bitcoin: Satoshi Nakamoto, "Bitcoin: A

Peer-to-Peer Electronic Cash System" (2008), https://bitcoin.org/bitcoin.pdf. Nostr protocol specification, https://github.com/nostr-protocol/nostr.

13
THE CRYPTO WARS

> "If privacy is outlawed, only outlaws will have privacy."
>
> Phil Zimmermann, "Why I Wrote PGP" (1991)[1]

The Crypto Wars are the ongoing conflict between states seeking surveillance capability and individuals developing privacy technology. The conflict began when strong cryptography moved from classified military research to civilian availability. States that had monopolized the most effective practical cryptography suddenly faced citizens with similar defensive capability. The response was predictable: attempts to control cryptographic technology and pressure systems toward backdoors.[2]

The largest attempts at broad control largely failed, for reasons economic analysis explains. But failure was not total, and the conflict continues.

13.1 History: Export Controls to Clipper Chip

Cryptography as Munitions

Until late 1996, many U.S. exports of strong cryptographic software were classified as munitions under the International Traffic in Arms Regulations (ITAR).[3] In late 1996, most commercial encryption items moved to Commerce controls, with further liberalization over the next several years. Sharing strong cryptographic software

across borders could trigger export-control penalties severe enough to chill researchers and developers.

The classification reflected Cold War assumptions: cryptography was military technology, and maintaining cryptographic superiority over adversaries justified restricting civilian access. That civilians might have legitimate privacy needs, independent of military considerations, did not factor into the regulatory framework.

The absurdity became apparent as computing proliferated. Mathematical formulas available in university libraries could trigger export problems when distributed as software. Regulators treated publication format and distribution channel as legally significant in ways courts later found constitutionally suspect: printed publication and academic discussion received stronger practical protection, while posting source code online could invite export-control conflict.

Phil Zimmermann and PGP

Phil Zimmermann's Pretty Good Privacy (PGP) crystallized the conflict.[1] In 1991, Zimmermann released PGP as free software, providing strong public-key cryptography to ordinary users. PGP spread rapidly via the early internet, soon reaching users outside the United States.

The federal government opened a criminal investigation. For three years, Zimmermann faced potential export-control prosecution. The case became a cause celebre in the nascent internet community. Zimmermann's response was characteristically cypherpunk: he published the PGP source code as a printed book, using a publication format that enjoyed stronger practical speech protection than export-controlled software files.[4]

The case was dropped without charges, but it established the template for Crypto Wars conflicts: the government asserts control authority; technologists route around restrictions; the restrictions prove unenforceable; formal policy catches up with technical reality.

The Clipper Chip

In 1993, the Clinton administration announced the Clipper Chip initiative: an NSA-designed cryptographic chipset with government-mandated key escrow.[5] Users would have encryption, but government agencies would hold duplicate keys enabling decryption when legally authorized.

The proposal combined encryption with surveillance. Proponents argued this balanced privacy against law enforcement needs. Opponents identified fundamental problems.

Escrowed keys create a single point of failure; if the escrow database were compromised, all protected communications would be exposed simultaneously. The scheme also required trusting government agencies to access keys only when legally authorized, and given revelations about warrantless surveillance, this trust was not warranted. Researcher Matt Blaze discovered a flaw in Clipper's protocol allowing users to disable escrow functionality while maintaining encryption.[6] The system designed to ensure government access could be bypassed in ways that undermined its core assurance. Finally, technology companies recognized that products with government backdoors would lose international markets; customers seeking actual privacy would choose products without mandated vulnerabilities.

Clipper was effectively abandoned as market rejection made it commercially unviable. The episode showed that mandated backdoors face both technical and economic obstacles.

Resolution of the 1990s Battles

By the late 1990s, the first Crypto Wars were winding down. Legal challenges and commercial pressure converged with practical unenforceability, pushing policy toward liberalization. In 1996, Executive Order 13026 began transferring encryption controls from the State Department to the Commerce Department. The Bernstein litigation helped establish source code as protected speech and weakened the regulatory framework further.[7]

Export controls were substantially relaxed, though not eliminated. Strong cryptography became legal to distribute and deploy. The infrastructure for encrypted communication that we now take for granted became possible.

13.2 The Economic Logic of Cryptographic Control

Why States Seek Control

States seek cryptographic control because encryption threatens surveillance capability. The reasons connect to core state functions.

States extract resources through taxation, and full financial surveillance enables tax enforcement; encrypted financial transactions, invisible to authorities, undermine enforcement capability. As Chapter 10 examined, monetary systems enable state control, while encrypted payment systems route around that control by enabling transactions outside monitored channels. States monitor populations for various purposes: identifying dissent, tracking movements, understanding social networks. Encryption creates spaces invisible to such monitoring. Investigating crimes often requires accessing communications and records, and encryption can prevent access even with legal authority.

From the state's perspective, encryption is a capability problem. Citizens with strong encryption can act without state visibility. This constrains state action regardless of whether that action is legitimate law enforcement or illegitimate repression.

Information Asymmetry and State Power

States benefit from information asymmetry: knowing more about citizens than citizens know about states. This asymmetry enables selective enforcement. It also chills behavior and enables preemptive intervention. When authorities can see all violations but must choose which to prosecute, enforcement becomes discretionary; everyone is guilty of something, and prosecution depends on official favor. Knowledge of surveillance changes behavior, as citizens who know they are watched modify actions to avoid attention,

even when those actions are legal. Early detection of organizing or resistance enables intervention before movements gain strength.

Encryption reduces information asymmetry. States see less; citizens can coordinate without visibility. The power that asymmetry provides is diminished.

Economic Stakes

The economic stakes are substantial on both sides. For states, surveillance infrastructure is a massive investment; entire agencies have built capabilities premised on access, and encryption threatens the return on that investment. For citizens, privacy enables economic activity that surveillance would prevent; underground economies, regulatory arbitrage, protection of competitive information, and simple preference for non-observed life all have value to those who want them. For businesses, the tension runs in both directions: governments demand access while customers demand privacy, and the commercial value of serving privacy-conscious customers conflicts with regulatory compliance.

Democracy and the Surveillance Ratchet

Hans-Hermann Hoppe's 2001 analysis of democratic and monarchic orders names a feature of democratic rule that bears on the history of surveillance legislation.[8] Democratic rulers are temporary caretakers of the apparatus instead of owners of its long-term capital value, and that structural difference produces higher time preference in democratic decision-making than in hereditary monarchic decision-making.

Democratic surveillance legislation has consistently traded present political safety against future civil-liberties cost. The USA PATRIOT Act passed the House of Representatives with 357 yeas in October 2001, forty-five days after September 11.[9] Each expansion of financial-surveillance authority under the Bank Secrecy Act's successor legislation has passed with overwhelming bipartisan support in a political environment where the immediate political cost of opposing the expansion was visible and the long-term cost

of supporting it was diffuse. The Chat Control, UK Online Safety Act, and eIDAS 2.0 battles in Europe track the same pattern. The political constituency for opposition is specialists who care about the architecture, while the political constituency for passage is whoever can frame the proposal as a response to a specific immediate concern.

The Hoppean reading is that democratic rulers discount future civil-liberties costs because they will not be in office to bear them. A more cynical reading is that they do not discount the future costs at all, because they expect the costs to accrue to opposition groups they will not belong to. Both readings produce the same pattern, and the pattern is what this book's cost-asymmetry argument contends with.

13.3 Why Control Fails (and Where It Doesn't)

The Structural Obstacles

Cryptographic security rests on mathematical properties that legal prohibition cannot change. If a problem is computationally hard, it remains hard regardless of what legislators decree. No law can make factoring large primes easy, and no policy reverses the relationships that make the schemes work. Cryptographic knowledge can be independently discovered, so suppressing it in one jurisdiction does not prevent discovery elsewhere; the mathematical relationships exist whether anyone knows them or not. Once published, mathematical knowledge cannot be unpublished; papers, textbooks, and internet archives preserve cryptographic techniques permanently and globally. Once algorithms are published, competent programmers can implement them directly, so suppressing implementations pushes toward suppressing programming itself.

The near-zero marginal cost of information compounds the problem. A cryptographic algorithm, once discovered and published, can be copied infinitely at negligible cost. One escaped copy becomes unlimited copies, and encryption software spreads faster than enforcement can track. The internet enables global distribution faster than any national enforcement mechanism can respond, and

software published in one jurisdiction is available worldwide within minutes.

Even ignoring the distributed-knowledge problem, effective control would require coordination among states with divergent interests. Not all states want to restrict encryption; some benefit economically from serving as havens for privacy technology development. Each state controls only its own territory, so a backdoor mandate in one country does not touch software developed elsewhere. Even coordinated international agreements face enforcement gaps, and motivated actors can find jurisdictions that neither participate nor enforce.

Where Control Succeeds

Despite these obstacles, cryptographic control is not entirely ineffective. It succeeds not against the mathematics but around it. Cryptography is hard to implement correctly, and most users cannot evaluate whether a given implementation is secure; this creates opportunities for compromised implementations to spread under benign-looking names. Secure systems often sacrifice usability, so when encryption is hard to use, people use it less or use it incorrectly, undermining security in practice while it remains sound in theory. Most users accept defaults, and systems that ship with weak or absent default encryption leave most users unprotected regardless of what strong options formally exist. Companies operating within a jurisdiction must comply with local law or face sanction; major platforms often implement surveillance capabilities because regulatory compliance requires it, not because their engineers prefer insecurity. And encryption protects data, not people: physical coercion can compel key disclosure regardless of cryptographic strength, and the "$5 wrench attack" remains effective against the uncareful target.

Control therefore fails against sophisticated, motivated actors operating with good tools and good practices. It often succeeds against ordinary users who lack the expertise or the motivation to implement strong privacy. The resulting two-tier situation, in which those who prioritize privacy get it while everyone else does not, is the equilibrium the rest of the book takes for granted.

13.4 Jurisdictional Competition and Arbitrage

Jurisdictions compete for economic activity, and technology development is part of that competition. Privacy-friendly jurisdictions attract both developers and users. Programmers prefer working where their work is legal, and encryption development has clustered in jurisdictions with favorable treatment. Businesses serving privacy-conscious customers locate where they can legally do so; jurisdictions such as Switzerland and Estonia have attracted privacy-focused technology companies. High-value users seeking privacy choose service providers partly by jurisdiction, and demand for offshore services reflects the same arbitrage dynamic operating on the consumption side.

Economist Charles Tiebout analyzed how competition among jurisdictions for residents creates pressure toward policies residents prefer.[10] Applied to cryptography, the analysis illuminates several dynamics at once. Developers and companies can relocate, and the credible threat of exit constrains what any single jurisdiction can impose. Jurisdictions known for privacy protection attract privacy-seeking activity, and the resulting reputation becomes an asset worth maintaining. When some jurisdictions offer favorable treatment, others face pressure to match or lose economic activity to those that do.

This competition can race toward privacy protection or toward surveillance, depending on which pressures a given government responds to. Governments responding primarily to law enforcement and intelligence pressures compete to offer more surveillance capability, racing toward the bottom. Governments responding primarily to economic development pressures compete to offer more privacy protection, racing toward the top. The outcome depends on which incentives dominate in which jurisdictions, and the present picture is mixed: some jurisdictions compete on privacy while others expand surveillance, and the same actors sometimes find themselves on opposite sides of the line as they cross borders.

13.5 The Ongoing War

The Shift from Tools to Builders

The Crypto Wars did not end with 1990s liberalization. They intensified, but along a different axis. The earlier battles tried to stop cryptography itself from escaping military control; once strong encryption became ordinary software, that front collapsed. States then adapted. If the mathematics could not be suppressed and the software could not be contained, the remaining pressure point was human.

A protocol cannot be arrested. A developer can. A piece of open-source code is legally awkward to prosecute directly; the person who wrote it, deployed it, promoted it, or operated infrastructure built around it is a clear legal target. Sanctions regimes, money-transmission statutes, anti-money-laundering law, and conspiracy doctrines all reach individuals regardless of what the underlying protocol does or does not do. The second phase of the Crypto Wars is a war against builders, because builders are the only remaining human chokepoint in systems designed to eliminate human chokepoints.

The chilling effect is the mechanism. When writing privacy-preserving code can carry the risk of a decade in prison under aggressive readings of sanctions or money-laundering law, fewer people write such code. When operating private-commerce infrastructure can be prosecuted as a criminal enterprise regardless of how users behave, fewer operators build. The prosecutions do not need to succeed on every legal theory to succeed as policy; they only need to raise the personal cost of building to a level that deters the marginal entrant.

Two cases anchor the pattern. Bernard von NotHaus's Liberty Dollar, which circulated private silver and gold pieces and paper certificates intended for use as current money alongside Federal Reserve notes from 1998 through the mid-2000s, drew a federal counterfeiting and conspiracy prosecution that ended the operation.[11] The theory was not that von NotHaus had operated an unlicensed financial service but that privately minted coins designed to circulate

as current money attacked the monetary unit itself. The jury instruction framed Liberty Dollar as an attempt to undermine public faith in United States currency. Physical centralized competition with state money, in identifiable hands, attracts a more direct legal answer than digital systems usually face.

Tornado Cash showed the digital version of the same pressure operating without a custodian to seize. Its core pool contracts became immutable on Ethereum once deployed, so no officer could turn them off. On August 8, 2022, the Office of Foreign Assets Control added Tornado Cash itself to the Specially Designated Nationals list, targeting the smart-contract code and the Ethereum addresses at which it ran instead of a person or a company; U.S. persons were prohibited from interacting with those addresses regardless of purpose.[12] The designation did not hold. In November 2024 the Fifth Circuit held in *Van Loon v. Department of the Treasury* that Tornado Cash's immutable smart contracts were not "property" under the International Emergency Economic Powers Act and could not be designated under OFAC's statutory authority, and on March 21, 2025, Treasury removed Tornado Cash from the sanctions list. The criminal cases against the developers who wrote the protocol, including Alexey Pertsev in the Netherlands and Roman Storm in the United States, continued on separate legal grounds and were unaffected by the delisting.[13] Protocol resilience and personal safety are different problems. Immutable code can outlast direct suppression, and the humans who wrote it absorb the legal pressure the protocol deflects. Control has migrated from the code to the coder, from the system to the people who maintain it.

The Commercial Spyware Arm

Encryption has won the wire. A message protected by a modern ratchet is unreadable in flight, so the state buys its way back in at the endpoint, where decrypted content is available by the same pathway the user sees it. Commercial spyware is the instrument the state reaches for once direct control of the math and prosecution of builders have failed to reach the content itself. Chapter 12 documents the industry's structure and vendor accountability record;

the Crypto Wars reading is narrower, placing spyware as the third phase after cryptographic control and builder prosecution have run their course.[14]

Content encryption, however mathematically secure, is insufficient once the endpoint is compromised. Chapter 22 takes up the operational response.

The Current Legal Battles

Proposals for "responsible encryption" with "exceptional access" continue to emerge with arguments close to what Clipper offered three decades earlier, and the technical problems remain. Since 2022, a cluster of specific legislative projects has made the current phase of the conflict unusually concrete.

The European Union's proposed Regulation on Child Sexual Abuse (CSAR), commonly called "Chat Control," would require providers of interpersonal communications services to scan user content for child sexual abuse material. The scope covers services that implement end-to-end encryption. Scanning end-to-end-encrypted content requires either breaking the encryption or inserting a scanner at the endpoint before encryption occurs; both paths amount to the same architectural compromise. Chat Control is triangular intervention applied to messaging: the state modifies the terms of the provider-user exchange by forcing the provider to surveil on the state's behalf, in the same way Chapter 10 described the state forcing banks to surveil their depositors. Germany, Austria, the Netherlands, and several other member states have blocked qualified majority across successive Council presidencies, and the European Parliament's LIBE Committee rejected mandatory scanning in its own November 2023 position. The text has been revised and reintroduced across four years without adoption. The durability of the opposition shows that the Clipper-era arguments about backdoor risk remain persuasive when they are presented as architectural claims instead of policy preferences.

The United Kingdom's Online Safety Act 2023 is further along. Section 122 of the Act grants the communications regulator Ofcom the power to issue notices requiring "accredited technology" to

scan end-to-end-encrypted communications for child sexual abuse material and for terrorism-related content. At the time of the Act's passage, the government issued a ministerial statement conceding that the power would not be exercised until "technically feasible" scanning became available. The statutory power remains, and its exercise depends on regulatory judgment, not on any additional primary legislation. Each of the major end-to-end-encrypted services operating in the UK, including Signal and WhatsApp and Apple's iMessage, announced that they would withdraw their services from the United Kingdom before implementing client-side scanning. The European Court of Human Rights held in *Podchasov v. Russia* in February 2024 that a statutory requirement to weaken end-to-end encryption "cannot be regarded as necessary in a democratic society," which is the first high-court holding in a Council of Europe member state that the architectural compromise fails a proportionality test.

The European Union's eIDAS 2.0 regulation introduces a different attack vector. Article 45 of the regulation, as originally drafted, would have compelled web browsers to trust Qualified Website Authentication Certificates issued by member-state-designated providers, without the browsers' usual authority to audit, revoke, or remove certificates they consider insecure. The mechanism would have enabled any designated member state to issue certificates that browsers are obliged to accept, which is the architectural definition of a man-in-the-middle capability against Transport Layer Security. Browser vendors including Mozilla, civil-liberties organizations including the Electronic Frontier Foundation, and more than five hundred academic cryptographers signed open letters against the draft. The final text, adopted in April 2024, includes a derogation clause permitting browsers to take "precautionary measures" when there are "substantiated concerns related to breaches of security"; browser vendors have interpreted this as preserving Certificate Transparency logging and revocation authority. The underlying mandate to trust qualified certificates remains, which is the compromise the regulation was designed to produce.[15]

The cypherpunk read of these three episodes is not that regulation always loses, because Ofcom's Section 122 powers are real

and can be exercised. Nor is it that regulation always wins, because Chat Control has been blocked for four years. The read is that the architectural arguments against backdoors, which were developed in response to Clipper, have survived three decades of technological change and now constitute the only durable defense against the current wave. When the argument is framed as "this is a backdoor," policy opposition holds. When the argument is framed as "this is a targeted mechanism for a serious crime," policy opposition weakens. The architectural frame is the one defenders must maintain.

The NeuralHash Withdrawal

In August 2021 Apple announced NeuralHash, a system that would have scanned images on user devices against a hash set of known child sexual abuse material before the images were uploaded to iCloud. The system was presented as privacy-preserving: the hash comparison would occur on the device, and only flagged matches would be reported. Within a month, security researchers published adversarial collisions against the NeuralHash function, showing that a determined attacker could craft images that produced false-positive matches against arbitrary target hashes. The objection from academic cryptographers, from civil-liberties organizations, and from Apple's own engineering community concerned architecture, with the false-positive rate as a secondary concern. A scanner on every device, even one scanning for a single content class on the day of deployment, was an infrastructure whose future uses would not be controlled by the provider.

Apple withdrew the system in December 2022. No successor has been deployed. The firm redirected its effort into an on-device, opt-in Communication Safety feature that does not report outside the device. The withdrawal showed that architectural critique can overturn a commercial deployment before the infrastructure is entrenched. The same critique applies to every client-side scanning proposal that follows, and the same cryptographic objections continue to apply.[16]

Continuing Regulatory Threats

Platform liability for user content creates incentives to surveil users and undermine end-to-end encryption. Cryptocurrency regulation extends financial surveillance through KYC requirements, exchange registration, and travel rules.[17] Proposals for international frameworks to govern encryption seek to close jurisdictional arbitrage opportunities.

The ``Going Dark'' Debate

Law enforcement agencies argue they are "going dark": losing access to communications that encryption protects, and the FBI has campaigned for mandatory access capabilities.[18] Access has in fact expanded, not contracted; despite encryption, law enforcement has more access to more data than ever before, with metadata, location tracking, and platform cooperation providing vast information streams. Every security expert who has examined the question concludes that mandated access introduces vulnerabilities that spread beyond police use and become targets for compromise by hostile actors.[19]

Post-Quantum Concerns

Quantum computing threatens current public-key cryptography. A sufficiently powerful quantum computer could break RSA and elliptic curve cryptography that secure most current internet traffic.[20][21]

This creates both threat and opportunity. Encrypted data captured today could be decrypted later when quantum computers mature, making "harvest now, decrypt later" a viable strategy for patient adversaries.[22] At the same time, post-quantum cryptography has moved from research into early standardization and deployment planning; NIST finalized its first core standards in 2024, but the transition to quantum-resistant algorithms remains a major infrastructure project.[23] States may try to use that transition as an opportunity to mandate backdoors in new cryptographic standards.

The Conflict Continues

The Crypto Wars continue as a durable feature of the relationship between states and citizens with access to strong cryptography. States want surveillance capability; citizens want privacy; cryptography can provide privacy that resists state surveillance; states therefore seek to constrain cryptography. Neither side can permanently win. Cryptography cannot be uninvented and state power cannot be abolished, and the conflict persists because both sides have durable interests.

Chapter Summary

Cryptographic control fails against sophisticated actors for structural reasons. Mathematics is indifferent to legal prohibition, information replication costs nearly nothing, and global coordination among jurisdictions with divergent interests is practically impossible. Control still succeeds against ordinary users through implementation difficulty and usability barriers: most users accept defaults, cannot evaluate cryptographic security, and operate within jurisdictions that compel institutional compliance. The two-tier equilibrium, strong protection for those who prioritize it, weak protection for everyone else, is the empirical result of three decades of conflict.

The current phase targets builders. Sanctions, money-transmission statutes, and conspiracy doctrines reach individuals in ways protocols cannot be reached, and prosecutions need not succeed on every theory to succeed as policy. A concrete cluster of regulatory initiatives carries the phase: the European Union's Chat Control proposal, the United Kingdom's Online Safety Act Section 122 powers, and the eIDAS 2.0 browser-trust mandate. Each has drawn on the Clipper-era vocabulary of "exceptional access" and has been resisted on architectural grounds, and Apple's withdrawal of the NeuralHash client-side scanning proposal in December 2022 showed that the critique can still overturn a commercial deployment before it becomes entrenched. Commercial spyware is the state's answer to the fact that content encryption has won the wire, and the spyware industry survives sanctions on individual vendors through

ordinary market substitution. Control has migrated from the code to the coder and from the transport to the endpoint, and Chapter 22 takes up the operational response.

The conflict is durable. Cryptography cannot be uninvented and state power cannot be abolished, and jurisdictional arbitrage opens opportunities without guarantees, and race dynamics can move toward surveillance as readily as toward privacy depending on which pressures governments respond to. Chapter 14 develops the cryptographic foundations this chapter assumes, and Part V examines the specific implementations the political and economic analysis here presupposes.[24]

Notes

[1]Phil Zimmermann, "Why I Wrote PGP," originally published in the PGP User's Guide (1991, updated 1999), `https://web.archive.org/web/20240101000000/https://www.philzimmermann.com/EN/essays/WhyIWrotePGP.html`. See also Chapter 2, note 15 for the cypherpunk-history treatment.

[2]The cypherpunk movement provided the intellectual and organizational context from which the Crypto Wars were fought on the civilian side. Two founding documents anchor it. Eric Hughes, "A Cypherpunk's Manifesto" (March 9, 1993), `https://www.activism.net/cypherpunk/manifesto.html`: "Privacy is necessary for an open society in the electronic age...We the Cypherpunks are dedicated to building anonymous systems." Timothy C. May, "The Crypto Anarchist Manifesto" (1988, distributed at the Hackers Conference; published to the Cypherpunks mailing list in 1992), `https://www.activism.net/cypherpunk/crypto-anarchy.html`, predicted that public-key cryptography would dissolve state control of information flows entirely. The Cypherpunks mailing list, active from 1992, is archived at `https://mailing-list-archive.cryptoanarchy.wiki/`. Steven Levy, *Crypto: How the Code Rebels Beat the Government — Saving Privacy in the Digital Age* (Viking, 2001), is the standard popular history of this period.

[3]For a fuller history of U.S. cryptographic export controls, see Whitfield Diffie and Susan Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption* (Cambridge, MA: MIT Press, 1998).

[4]The printed source code strategy exploited a legal distinction in practice. Regulators treated software as export-regulated, while books received stronger and clearer speech protection. Philip Zimmermann, *PGP: Source Code and Internals* (Cambridge, MA: MIT Press, 1995).

[5]On the Clipper Chip proposal and controversy, see A. Michael Froomkin, "The Metaphor Is the Key: Cryptography, the Clipper Chip, and the Constitution," *University of Pennsylvania Law Review* 143, no. 3 (1995): 709–897.

[6]Matt Blaze, "Protocol Failure in the Escrowed Encryption Standard," *Proceedings of the 2nd ACM Conference on Computer and Communications Security* (1994): 59–67.

[7]*Bernstein v. U.S. Department of State*, 922 F. Supp. 1426 (N.D. Cal. 1996), and the subsequent Ninth Circuit panel opinion of May 6, 1999, *Bernstein v. U.S. Department of Justice*, 176 F.3d 1132 (9th Cir. 1999), holding that software source code was protected speech under the First Amendment. The opinion was withdrawn when the case went en banc, but the litigation contributed to the parallel liberalization of export controls under Executive Order 13026 (November 15, 1996) and subsequent Bureau of Industry and Security rule revisions.

[8]Hans-Hermann Hoppe, *Democracy: The God That Failed: The Economics and Politics of Monarchy, Democracy, and Natural Order* (New Brunswick, NJ: Transaction Publishers, 2001). The time-preference argument is most concentrated in chapter 1 ("On Time Preference, Government, and the Process of Decivilization") and chapter 2. Austrian-internal criticism includes Joseph T. Salerno, "Comment on Hoppe's 'On Time Preference,"' *Review of Austrian Economics* 9, no. 1 (1996), and later exchanges; Walter Block, "Is Monarchy a Libertarian Stance?" *Reason Papers* 32 (2010), defends Hoppe against some objections. The framework is used in the main text as a lens on surveillance legislation, not as a general endorsement of Hoppe's monarchic-versus-democratic comparison.

[9]Roll call vote no. 398, Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism Act of 2001 (USA PATRIOT Act, H.R. 3162), 107th Congress, October 24, 2001: 357 yeas, 66 nays. Office of the Clerk of the U.S. House of Representatives, `https://clerk.house.gov/Votes/2001398`. The Senate passed the bill the following day by 98-1.

[10]Charles M. Tiebout, "A Pure Theory of Local Expenditures," *Journal of Political Economy* 64, no. 5 (1956): 416–424.

[11]Bernard von NotHaus founded the Liberty Dollar in 1998 as a private precious-metal currency project; the federal prosecution centered on counterfeiting and conspiracy theories involving privately minted coins intended for use as current money. The conviction was upheld on appeal and Liberty Dollar's operations ended thereafter. See *United States v. Bernard von NotHaus*, W.D.N.C. 5:09-cr-00027 (2011), and the Fourth Circuit's affirmance.

[12]U.S. Department of the Treasury, Office of Foreign Assets Control, designation of Tornado Cash, August 8, 2022, `https://home.treasury.gov/news/press-releases/jy0916`, listing numerous associated smart-contract addresses. Criminal charges against developers followed, including Alexey Pertsev in the Netherlands (convicted May 2024) and Roman Storm in the United States.

[13]*Van Loon v. Department of the Treasury*, 122 F.4th 549 (5th Cir. 2024), held that Tornado Cash's immutable smart contracts were not "property" under the

International Emergency Economic Powers Act and could not be designated under OFAC's statutory authority. Treasury formally removed Tornado Cash from the Specially Designated Nationals list on March 21, 2025, citing the "novel legal and policy issues raised by use of financial sanctions against financial and commercial activity occurring within evolving technology and legal environments." Criminal cases against developers continued on separate legal grounds.

[14]The commercial spyware industry is mature, globally deployed, and resistant to sanctions on individual vendors; see Chapter 12 note 8 for the full product-and-legal stack (NSO Pegasus, Intellexa Predator, Paragon Graphite, Candiru, Cellebrite, Magnet Forensics, and the U.S. NSO ruling). Citizen Lab, `https://citizenlab.ca/`, is the primary empirical source tracking deployments across state customers. *WhatsApp LLC v. NSO Group*, Case No. 4:19-cv-07123 (N.D. Cal. 2024), the first major civil accountability ruling against a spyware vendor. NSO Group, `https://www.nsogroup.com/`. U.S. Commerce Department Entity List designation of NSO Group (November 3, 2021), 86 Fed. Reg. 60759 (November 4, 2021).

[15]Chat Control (EU CSAR proposal), the UK Online Safety Act Section 122, and eIDAS 2.0 Article 45 form the current cluster of regulatory pressure against end-to-end encryption; the European Court of Human Rights ruled in Podchasov v. Russia (February 2024) that weakening E2EE is not necessary in a democratic society. EU Regulation on Child Sexual Abuse ("Chat Control" / CSAR), proposal COM(2022) 209, `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0209`, with European Parliament LIBE Committee position tracked at `https://www.europarl.europa.eu/committees/en/libe/home`. UK Online Safety Act 2023, Section 122, `https://www.legislation.gov.uk/ukpga/2023/50/section/122`. EU eIDAS 2.0 Regulation (EU) 2024/1183, `https://eur-lex.europa.eu/eli/reg/2024/1183/oj`. Mozilla, EFF, and cybersecurity-experts open letter on eIDAS 2.0 Article 45 at `https://blog.mozilla.org/en/mozilla/mozilla-eff-cybersecurity-experts-publish-letter-on-dangers-of-article-452-eidas-regulation/`. European Court of Human Rights, *Podchasov v. Russia* (App. no. 33696/19), judgment of February 13, 2024, `https://hudoc.echr.coe.int/eng?i=001-230854`. Signal Foundation, `https://signal.org/`, with public statements on withdrawing from jurisdictions that mandate client-side scanning archived at `https://signal.org/blog/`; see also Chapter 5, note 7.

[16]Apple announced NeuralHash in August 2021 as an on-device CSAM scanner, researchers published adversarial collisions within a month, and Apple withdrew the system in December 2022. Apple's original NeuralHash announcement (August 5, 2021), archived at `https://web.archive.org/web/20210805170250/https://www.apple.com/child-safety/`. Adversarial collision demonstration: AsuharietYgvar et al., "AppleNeuralHash2ONNX," GitHub repository,

https://github.com/AsuharietYgvar/AppleNeuralHash2ONNX. Academic and civil-liberties response: Harold Abelson et al., "Bugs in our Pockets: The Risks of Client-Side Scanning," October 2021, https://arxiv.org/abs/2110.07450. Apple's current on-device Communication Safety feature documented at https://support.apple.com/guide/iphone/communication-safety-iph00618be7a5/ios.

[17]The Financial Action Task Force Travel Rule (FATF Recommendation 16) requires virtual asset service providers to pass originator and beneficiary information with transfers above a threshold, extending bank-secrecy-style KYC obligations to cryptocurrency exchanges. FATF, *Updated Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers* (October 2021), https://www.fatf-gafi.org/publications/fatfrecommendations/documents/guidance-rba-virtual-assets-2021.html. Domestic implementation in the United States proceeds through FinCEN under the Bank Secrecy Act; the proposed Travel Rule rulemaking for cryptocurrency is tracked at https://www.fincen.gov/. The EU's Transfer of Funds Regulation (EU) 2023/1113 extended the Travel Rule to crypto-asset service providers from December 30, 2024. These requirements replicate the financial-surveillance architecture Chapter 10 analyzed: conditioning service access on disclosure, applied across the cryptocurrency layer.

[18]James Comey, "Going Dark: Are Technology, Privacy, and Public Safety on a Collision Course?" speech at the Brookings Institution, October 16, 2014, https://www.fbi.gov/news/speeches/going-dark-are-technology-privacy-and-public-safety-on-a-collision-course.

[19]Harold Abelson et al., "Keys Under Doormats: Mandating Insecurity by Requiring Government Access to All Data and Communications," *Journal of Cybersecurity* 1, no. 1 (2015): 69–79.

[20]Daniel J. Bernstein and Tanja Lange, "Post-Quantum Cryptography," *Nature* 549 (2017): 188–194.

[21]The mathematical foundations of public-key cryptography that the chapter identifies as legally indifferent to prohibition originate in two landmark papers. Whitfield Diffie and Martin E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* 22, no. 6 (1976): 644–654, introduced the Diffie-Hellman key exchange and the concept of asymmetric cryptography. Ronald L. Rivest, Adi Shamir, and Leonard Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM* 21, no. 2 (1978): 120–126, introduced RSA. Neal Koblitz, "Elliptic Curve Cryptosystems," *Mathematics of Computation* 48, no. 177 (1987): 203–209, and Victor S. Miller, "Use of Elliptic Curves in Cryptography," in *Advances in Cryptology — CRYPTO '85*, Lecture Notes in Computer Science 218 (Springer, 1986): 417–426, introduced elliptic-curve cryptography. All four are the mathematical relationships the text argues cannot be reversed by legislative decree.

[22]The "harvest now, decrypt later" threat describes adversary programs that capture and store currently encrypted traffic against the day a cryptanalytic

capability makes retrospective decryption feasible. The NSA's MUSCULAR and BULLRUN programs, revealed through the Snowden documents in 2013, showed that bulk interception of encrypted traffic was already operational policy. Glenn Greenwald and Ewen MacAskill, "NSA Prism Program Taps into User Data of Apple, Google and Others," *The Guardian* (June 6, 2013); James Ball, Julian Borger, and Glenn Greenwald, "Revealed: How US and UK Spy Agencies Defeat Internet Privacy and Security," *The Guardian* (September 5, 2013), on BULLRUN/EDGEHILL. NIST's post-quantum migration guidance explicitly names harvest-now-decrypt-later as a primary threat motivating urgency in the transition: NIST, *Migration to Post-Quantum Cryptography*, Special Publication 1800-38 (2023), `https://csrc.nist.gov/publications/detail/sp/1800-38/final`.

[23]NIST finalized FIPS 203 (ML-KEM), FIPS 204 (ML-DSA), and FIPS 205 (SLH-DSA) on August 13, 2024, `https://csrc.nist.gov/projects/post-quantum-cryptography`. See also Chapter 5, note 4 for the broader NIST PQC standardization context.

[24]Further reading on state surveillance history. For the pre-9/11 history of U.S. mass surveillance, the canonical text is James Bamford's trilogy: *The Puzzle Palace: A Report on America's Most Secret Agency* (Houghton Mifflin, 1982), *Body of Secrets* (Doubleday, 2001), and *The Shadow Factory: The Ultra-Secret NSA from 9/11 to the Eavesdropping on America* (Doubleday, 2008). For the post-2001 surveillance expansion, Dana Priest and William Arkin, *Top Secret America: The Rise of the New American Security State* (Little, Brown, 2011), is the standard investigative treatment. On the Snowden revelations specifically, Glenn Greenwald, *No Place to Hide* (Metropolitan Books, 2014), is the journalist's account; Luke Harding, *The Snowden Files* (Vintage, 2014), covers the *Guardian*'s side; Edward Snowden, *Permanent Record* (Metropolitan Books, 2019), is Snowden's memoir. The unpublished Church Committee reports, *Intelligence Activities and the Rights of Americans*, Senate Report No. 94-755 (1976), remain the broadest congressional investigation of U.S. surveillance and are available at `https://www.intelligence.senate.gov/resources/intelligence-related-commissions`. For the FBI side, Athan Theoharis, *Spying on Americans: Political Surveillance from Hoover to the Huston Plan* (Temple University Press, 1978), and Tim Weiner, *Enemies: A History of the FBI* (Random House, 2012), are the two standard histories. On the broader technical infrastructure, Whitfield Diffie and Susan Landau, *Privacy on the Line*, already cited, and Keith Devlin, *The Math Behind the Scenes* (Springer, 2022), cover the cryptographic-policy history. For contemporary mass-surveillance practice outside the U.S., Ronald Deibert, *Reset: Reclaiming the Internet for Civil Society* (House of Anansi, 2020), is the Citizen Lab director's synthesis; Cédric Biscay et al., *Privacy International's State of Privacy* reports at `https://privacyinternational.org`, track specific programs internationally. On the doctrinal and constitutional sides, Daniel J. Solove and Paul M. Schwartz, *Information Privacy Law*, 7th ed. (Aspen, 2020), is the standard U.S. legal treatise; Orin S. Kerr, *The Digital Fourth Amendment* (Oxford University

Press, 2024), is the current reference on Fourth Amendment doctrine in digital contexts.

PART V

TECHNICAL IMPLEMENTATION

14

CRYPTOGRAPHIC FOUNDATIONS

"The universe believes in encryption. It is easier to encrypt information than it is to decrypt it."

Julian Assange, *Cypherpunks: Freedom and the Future of the Internet* (2012)[1]

The goal of this chapter is conceptual understanding, not a full cryptography course. Readers need to understand what cryptographic tools accomplish and why they work, not how to implement them. Implementation requires specialized expertise; using implementations requires understanding their properties. Readers who want to go deeper should consult David Wong, *Real-World Cryptography* (Manning, 2021) for an accessible applied introduction, Jonathan Katz and Yehuda Lindell, *Introduction to Modern Cryptography*, 3rd ed. (CRC Press, 2020) for rigorous formal treatment, Justin Thaler, *Proofs, Arguments, and Zero-Knowledge* (2022, available free online) for the proof-systems side, and Ross Anderson, *Security Engineering*, 3rd ed. (Wiley, 2020) for how cryptography interacts with everything around it.[2] The foundational papers of the discipline, its political and ethical framing, and its longer historical narrative are collected in a separate further-reading note at the end of the chapter.[3]

Cryptography solves coordination problems through mathematics, not institutions. Where traditional systems require trusting intermediaries, cryptographic systems require trusting only com-

putational hardness assumptions. This shift from institutional to mathematical trust is the foundation for privacy-preserving technology.

Three jobs must stay separate from the beginning. One tool hides content, another proves who authorized a message, and a third lets strangers verify integrity without trusting the same intermediary. The chapter moves through those jobs in that order, then asks what mathematics can and cannot replace.

14.1 Symmetric and Asymmetric Cryptography

Symmetric Cryptography

Symmetric cryptography is the oldest form of the first job: hiding content. One key encrypts; the same key decrypts. If only Alice and Bob possess the key, only they can read the message.

The speed of symmetric algorithms makes them the workhorse of content protection. Modern ciphers (AES, ChaCha20) encrypt data at gigabytes per second.[4] But speed alone does not solve the trust problem.

How do Alice and Bob establish a shared secret without meeting in person? If they communicate the key over an insecure channel, an eavesdropper intercepts it. If they need a secure channel to exchange the key, they already have secure communication and do not need the key. This is the key distribution problem, and for millennia it confined cryptography to parties who could physically exchange secrets: diplomats with couriers, military with secure channels, spies with dead drops. Mass adoption of encryption required breaking that confinement.

Asymmetric Cryptography

Asymmetric (public-key) cryptography, discovered in the 1970s, broke the key-distribution bottleneck.[5] Instead of one shared key, each party generates a mathematically related key pair: a public key they can share openly and a private key they keep secret.

The mathematical relationship between the two keys is non-reversible: computing the public key from the private key is simple, but computing the private key from the public key is computationally infeasible. No prior relationship is required. Alice can send Bob an encrypted message having never communicated with him before, using only his publicly available public key. The trust requirement shifts from "do I have a secure channel to this person?" to "is this public key authentic?" That second question is solvable without institutions.

This resolves the confidentiality side of key distribution. Alice publishes her public key; others encrypt messages to her; only she can decrypt them. No secure channel is needed, and no intermediary. Authenticity of the public key remains a separate problem, addressed below.

The Hybrid Reality

Symmetric cryptography hides content efficiently when parties already share a secret. Asymmetric cryptography eliminates the need for that prior secret. Neither alone suffices.

In practice, systems use both. Asymmetric cryptography establishes a session key; symmetric cryptography encrypts the actual data. The hybrid approach means every encrypted connection on the internet rests on the same foundation: mathematical trust replacing institutional trust at the key-exchange layer, then raw speed protecting the content itself.

The Algorithms

Several foundational algorithms enable asymmetric cryptography. Diffie-Hellman key exchange, published in 1976, allows two parties to establish a shared secret over a public channel.[5] Neither party reveals their private key, but both derive the same shared secret through mathematical operations on public values. Diffie-Hellman solves key exchange but not encryption directly; the shared secret it produces typically becomes the key for symmetric encryption.

RSA, published in 1978, provides both encryption and digital signatures using the difficulty of factoring large prime numbers.[6] Security depends on factorization remaining computationally infeasible for sufficiently large numbers. RSA can encrypt messages directly (up to a size limit) and create signatures. Its disadvantage is key size: secure RSA requires keys of 2048 bits or more, making it slower and more resource-intensive than alternatives.

Elliptic Curve Cryptography (ECC), introduced in the mid-1980s, achieves equivalent security with smaller keys using different mathematical structures.[7] A 256-bit elliptic curve key provides security comparable to a 3072-bit RSA key. The smaller keys make ECC faster and more suitable for constrained devices. Bitcoin uses the secp256k1 elliptic curve for its signatures.[8] Most modern systems prefer ECC over RSA for new implementations.

In practice, these algorithms serve complementary roles. Diffie-Hellman (or its elliptic curve variant ECDH) establishes shared secrets; RSA or elliptic curve signatures authenticate parties; and the resulting shared secrets key symmetric ciphers like AES for bulk encryption.

The Role of Randomness

Cryptographic security depends on unpredictability. Keys must be randomly generated; if an attacker can guess or predict a key, the strongest algorithm provides no protection.

Entropy measures unpredictability. A 256-bit key has 256 bits of entropy only if each bit is equally likely to be 0 or 1, independent of all other bits. If key generation has bias or patterns, effective entropy falls below what the bit length suggests, leaving the key weaker than it appears.

Randomness in cryptography must be cryptographically secure: not just "random looking" but unpredictable to any adversary. A pseudorandom number generator (PRNG) is an algorithm that uses a small initial value called a seed to produce a long sequence of numbers that appear random but are deterministic. PRNGs that produce statistically random output may still be predictable if an attacker knows the internal state or seed value. Cryptographically

secure pseudorandom number generators (CSPRNGs) are designed so that even observing their output does not reveal future values.

Sources of entropy include hardware random number generators that sample physical phenomena (thermal noise, radioactive decay, electronic noise) and system events (keystroke timing, mouse movements, network packet arrival times). Since any single source might be compromised or insufficient, secure systems concatenate multiple independent entropy sources. An attacker who can predict one source still cannot predict the combined output if other sources remain unpredictable. Operating systems maintain entropy pools that accumulate randomness from all available sources and feed CSPRNGs that applications use for key generation.

When randomness fails, cryptography fails completely. The algorithms may be sound, but predictable keys are guessable keys.

14.2 Hash Functions and Digital Signatures

One-Way Functions

A cryptographic hash function takes input of any size and produces a fixed-size output (the "hash" or "digest"). SHA-256, widely used in Bitcoin and elsewhere, produces a 256-bit output regardless of input size.[9]

Hash functions exhibit several properties that together make them useful. The same input always produces the same output. Given the output, finding any input that produces it is computationally infeasible, and finding two different inputs that produce the same output is also computationally infeasible. They also exhibit the avalanche effect: small changes in input produce dramatically different outputs.

Hash functions enable efficient integrity verification. Instead of comparing entire files, compare their hashes. Matching hashes indicate matching files with overwhelming probability; differing hashes show the files differ.

Digital Signatures

Digital signatures use asymmetric cryptography to provide authentication and integrity.[10] Unlike encryption (where anyone with the public key encrypts and only the private key holder decrypts), signatures work in the opposite direction: only the private key holder can create a signature, but anyone with the public key can verify it.

The signing process begins by computing the hash of the document, creating a fixed-size digest of the content. The signer then applies the signature algorithm using their private key and the hash, producing a signature value that accompanies the document.

Verification reverses this process. The verifier independently hashes the document, then applies the verification algorithm using the public key, the signature, and the recomputed hash. The algorithm outputs valid or invalid.

The mathematics varies by scheme. RSA signatures involve modular exponentiation; ECDSA,[11] the scheme used in Bitcoin, involves elliptic curve point multiplication and modular arithmetic; and Schnorr signatures[12] use a different construction with useful algebraic properties. What they share is the core asymmetry: creating a valid signature requires the private key; verifying requires only the public key.

Signatures prove three things. Authentication: only someone with the private key could have created the signature, so if you trust the public key belongs to Alice, the signature proves Alice signed. Integrity: any modification to the document after signing invalidates the signature because the recomputed hash will not match. Non-repudiation: Alice cannot credibly deny having signed if the signature validates against her public key.

Trustless Verification

What does it mean to verify without trusting the verifier? Anyone with the public key can independently confirm a signature, so no authority needs to be asked and no intermediary can falsely claim the verification happened. The mathematics is self-enforcing.

"Trustless" has a narrow meaning here: the verification process itself requires no trust because anyone can perform it. The question of whether the public key belongs to its purported owner remains, but that is a different problem addressed below.

14.3 Trust: Mathematical vs. Institutional

Traditional Trust Models

Before cryptography, trust required institutions. Reputation allowed parties to build track records over time, though new entrants faced high barriers. Legal enforcement punished breach of agreements, but effectiveness depended on jurisdiction and resources. Trusted third parties served as intermediaries who vouched for unknown parties, concentrating trust in those intermediaries. Physical security through vaults or guards, along with sealed documents, provided tangible protection.

Each model has failure modes: reputation can be manufactured, enforcement requires access to legal systems, intermediaries can be corrupted or coerced, and physical security can be breached.

Digital systems inherited this pattern without changing it. To send email, share a file, or join a chatroom, you trusted a server administrator to keep your messages confidential and to enforce who could access them. Access control was a server-side policy, implemented as a database row or a permission flag or a configuration setting, running on infrastructure you did not control and could not inspect. If the administrator was honest, competent, and uncoerced, the system worked as advertised. If the administrator was malicious, subpoenaed, or breached, access policies were irrelevant and the data was exposed. The institutional trust problem had been translated into a system administration problem, and the trust requirement was no smaller for the translation.

Mathematical Trust

Cryptographic trust rests on computational hardness assumptions. The factorization assumption holds that factoring the product

of two large primes is computationally infeasible. The discrete logarithm assumption holds that computing discrete logarithms in certain groups is computationally infeasible. Hash function assumptions hold that finding collisions or preimages for properly designed hash functions is computationally infeasible. These assumptions have been studied for decades by mathematicians and cryptographers worldwide. Unlike institutional trust, they do not vary with personnel changes, political pressures, or economic incentives. Mathematics does not accept bribes.

End-to-end encryption changes where access rules are enforced. Instead of a server deciding who may read a message, the cryptography decides: only holders of the correct keys can decrypt, and no configuration change on the server can override that fact. Access control moves from a policy in the infrastructure to a property of the data itself. The administrator who was once the guarantor becomes irrelevant to the guarantee. Infrastructure continues to route and store the ciphertext, but it does so without being able to read it, and the trust the user must extend collapses from trust in people and institutions to trust in mathematics and in the key management practiced at the endpoints.

Why Mathematics Is More Reliable

Why should anyone trust a mathematical proof over a human institution? Because mathematical trust has properties institutional trust structurally lacks. The same proof verifies the same way everywhere; a valid signature in one country is valid in all countries. The algorithms are public; anyone can verify the mathematics; security does not depend on secrecy of method. Verification requires no third party. Alice can verify Bob's signature without asking anyone's permission or trusting any intermediary. And computational verification scales with hardware, while human verification does not.

Limits of Mathematical Trust

Mathematical trust is not unlimited. Mathematics cannot tell you whether a public key belongs to whom it claims; that requires external verification: meeting in person, a web of trust, certificate authorities, or some other out-of-band confirmation. The mathematics may be sound while the implementation is flawed, and software bugs can undermine theoretically perfect cryptography. Computational assumptions themselves could fail if P=NP or if quantum computers mature sufficiently. Users can also be tricked into revealing keys or trusting the wrong public keys, and they may rely on compromised software.

Mathematical trust replaces some trust requirements but not all, shifting reliance away from institutions and toward assumptions tested in code and mathematics. The shift is valuable but not absolute.

14.4 Limitations and Vulnerabilities

Implementation Bugs vs. Cryptographic Breaks

Cryptographic algorithms are rarely broken mathematically; what fails is implementation.

Buffer overflows allow attackers to overwrite memory and extract keys, while timing attacks measure how long operations take to reveal information about keys from the timing alone. Random number failures compromise security because cryptography requires unpredictable randomness. Protocol errors occur when individual algorithms are secure but their combination is not. Most real-world cryptographic failures are implementation failures. The mathematics can hold while the code fails.

Side-Channel Attacks

Side-channel attacks extract information from physical implementation, not from mathematical weakness. Power analysis measures power consumption during cryptographic operations to reveal key bits, and electromagnetic emanations from computing equip-

ment can leak information via radio signals. Cache timing attacks observe cache behavior to reveal memory access patterns correlated with keys. Acoustic attacks analyze sound produced by computers to leak cryptographic information. These attacks require physical proximity or sophisticated equipment but demonstrate that cryptographic security depends on more than algorithm strength.

The Human Element

Humans are the weakest link. Social engineering that convinces people to reveal keys or install malware bypasses cryptography entirely. Encryption protected by weak passwords provides weak protection. Key management presents persistent challenges: lost keys mean lost data, and compromised keys mean compromised data. Systems that are hard to use correctly are used incorrectly, and users disable security features that interfere with tasks.

Physical coercion, the "$5 wrench attack" examined in Chapter 5, remains outside cryptography's domain.[13] Cryptography protects data, not people.

The Quantum Horizon

Quantum computers threaten current public-key cryptography. Shor's algorithm, running on a sufficiently powerful quantum computer, could break RSA and elliptic curve cryptography by efficiently solving the mathematical problems they rely on.[14] The threat is not hypothetical in the academic sense, because the algorithms are published and their operation is well understood. The threat is hypothetical in the hardware sense, because the machines that would execute them at the required scale do not yet exist. Estimates of the year by which such a machine might appear vary across national-security agencies and industrial research programs, with the mainstream published range falling between 2030 and 2040. The defender's planning horizon must be calibrated to the earliest credible date, not the median.

Status varies by cryptographic type. Asymmetric cryptography, including RSA and ECC, is vulnerable to quantum attack through

Shor's algorithm. Symmetric cryptography is less affected; Grover's algorithm[15] provides only quadratic speedup, so doubling key lengths (for example, using AES-256 instead of AES-128) maintains security. Hash functions are similarly less affected, with quantum computers providing modest speedup that does not break them. Asymmetric primitives must be migrated; symmetric primitives need parameter adjustments but not architectural replacement.

The more pressing near-term threat is not live attack but "harvest now, decrypt later." An adversary who captures encrypted traffic today and retains it until quantum decryption becomes available can read that traffic when the machine arrives. The defender's exposure therefore depends on the longevity of the secret, not on the current state of quantum hardware. Traffic that must remain confidential for a decade or more should be assumed to be exposed to future quantum decryption even when captured today.

The Standards Arrive

NIST finalized three core standards in August 2024, ML-KEM for key exchange, ML-DSA for signatures, and SLH-DSA as a conservative hash-based backup, and added a code-based HQC backup in March 2025.[16] The standardization process took eight years of public cryptanalysis.

The transition from standardization to deployment has moved faster than any previous cryptographic transition. Signal deployed PQXDH (a hybrid combining X25519 and Kyber) in September 2023 before the NIST standard was finalized. Apple shipped iMessage PQ3 in iOS 17.4 and macOS 14.4 in February 2024. Google enabled hybrid Kyber-based key exchange by default in Chrome 124 in April 2024 and migrated to the standardized X25519MLKEM768 later that year. Cloudflare reported by mid-2025 that more than thirty percent of Transport Layer Security handshakes at its edge used hybrid post-quantum key exchange. The deployment pattern is consistent: mix the new primitive with a proven classical primitive so a break in either one does not compromise the session, and migrate the entire internet before the threat materializes.

Governments have published migration timelines on the same premise. The U.S. National Security Agency's CNSA 2.0 suite requires ML-KEM and ML-DSA for National Security Systems by 2033, with earlier adoption permitted where possible. The U.K. National Cyber Security Centre recommends post-quantum migration for high-value systems by 2035. Both timelines assume a cryptanalytically relevant quantum computer could appear before 2035, and both use 2033 to 2035 as the deadline by which migration must be complete, not as the expected appearance date. The principle is standard in security engineering: the migration budget is the gap between the deadline and the threat, and the deadline must precede the threat by a margin that accounts for implementation lag.

The Quantum Horizon and the Axiom of Resistance

The Axiom of Resistance holds that a system's security is measured by the cost required to compromise it. Post-quantum cryptography is an engineering response to a predictable change in that cost. Mathematical problems that were hard for classical computers become tractable for quantum computers; the cost of breaking them collapses. The axiom does not require that cost be infinite. It requires that the defender's cost of maintenance remain below the adversary's cost of compromise, and the post-quantum transition is what keeps that inequality in place across the hardware transition.

The hash-based construction illustrated by FIPS 205 depends only on the properties of hash functions, which are well understood and less affected by quantum computing. The primitive is conservative in exactly the sense the axiom calls for: when the assumption set is smaller, the failure surface is smaller. The lattice-based ML-KEM and ML-DSA standards rely on assumptions that are more recent and less studied than the assumptions underlying RSA. This is the cost of the transition. The defender accepts a less-tested assumption in exchange for quantum resistance, and the standards community's response has been to publish multiple independent algorithm families so that an unexpected attack on one family does not break the entire transition.

The harvest-now-decrypt-later threat model also forces a calibration of what qualifies as a long-lived secret. Conversations that were assumed ephemeral because they were encrypted in transit are long-lived if the ciphertext is captured. The defender who encrypted a message in 2024 under pre-quantum primitives must assume that the message will be readable at the point in the future when the machine arrives. This is why Signal and Apple and Chrome deployed their hybrid constructions before the threat materialized. The clock on past traffic cannot be rewound, but the clock on future traffic is still running.

The transition is a major infrastructure project. It is also technically feasible, and the deployments cited above demonstrate the feasibility at internet scale. Whether the defender completes the transition before the adversary benefits from captured pre-transition traffic is the variable that determines whether the harvest-now-decrypt-later threat lands.

What Cryptography Cannot Solve

Cryptography cannot solve endpoint security; if the device is compromised, cryptography on that device is meaningless. It cannot hide metadata; encryption hides content but not the fact of communication, and who talks to whom remains visible, along with timing and frequency, without additional protection (see Chapter 17). Physical coercion can still compel key disclosure. Cryptography also cannot solve social problems or make people trustworthy, only make certain betrayals detectable. Nor can it establish key authenticity, because mathematics cannot tell you if the public key belongs to its purported owner.

Cryptography is a tool. It solves specific problems. Expecting it to solve problems beyond its scope leads to false confidence.

Chapter Summary

Cryptography shifts trust from institutions to mathematics. Where traditional systems require trusting intermediaries, cryptographic systems require trusting only computational hardness

assumptions, the same body of assumptions tested by decades of failed attacks. Symmetric cryptography hides content efficiently when keys are shared; asymmetric cryptography eliminates the need for a prior shared secret by giving each party a key pair whose private half cannot be derived from the public half. Hybrid systems use both: asymmetric for key agreement, symmetric for bulk encryption. Hash functions produce fixed-size fingerprints that enable integrity verification, and digital signatures combine hashing with asymmetric cryptography to provide authentication, integrity, and non-repudiation. Anyone with the public key can verify a signature independently; no authority need confirm, and no intermediary can falsify the result.

Mathematical trust replaces some trust requirements but not all. Key authenticity still requires external verification, meeting in person, a web of trust, certificate authorities, or some out-of-band channel, because mathematics cannot tell whether a public key belongs to whom it claims. Implementations can be flawed even when algorithms are sound, and most real-world cryptographic failures originate in code: bad parameters, weak randomness, poor key handling, side channels. Side-channel attacks extract information from physical observables (power consumption, timing, electromagnetic emanations) while the underlying mathematics holds. Humans remain the weakest link: social engineering bypasses the mathematics entirely, and physical coercion compels disclosure regardless of key strength. Cryptography protects data, not people.

The quantum horizon has moved from research to production. NIST finalized the core lattice and hash-based signature standards (ML-KEM, ML-DSA, SLH-DSA) in August 2024 and selected HQC as a code-based backup KEM in March 2025. Signal's PQXDH protocol, Apple iMessage PQ3, Chrome's hybrid key exchange, and Cloudflare's edge all ship post-quantum cryptography in deployed consumer systems. The transition is a major infrastructure project that must complete before cryptanalytically relevant quantum computers arrive, because harvest-now-decrypt-later makes today's ciphertext an asset for tomorrow's decryption. Cryptography solves confidentiality, authentication, and integrity; it does not solve end-

point compromise, metadata exposure, physical coercion, or key authenticity, and the rest of the book addresses what cryptography alone cannot.

Notes

[1]Julian Assange, with Jacob Appelbaum, Andy Müller-Maguhn, and Jérémie Zimmermann, *Cypherpunks: Freedom and the Future of the Internet* (New York: OR Books, 2012). The quoted passage, from the opening of Assange's introduction, states the structural claim this chapter develops: that the asymmetry between the cost of encrypting information and the cost of decrypting it is a law of the universe, not a policy choice; strong cryptography therefore constrains state power architecturally, at the level of physics and computation, where law cannot reach. For the earlier cypherpunk-movement statement of the same goal in manifesto form, see Eric Hughes, "A Cypherpunk's Manifesto" (1993), `https://www.activism.net/cypherpunk/manifesto.html`.

[2]For open-source implementations readers can inspect and use, GnuPG (`https://gnupg.org`) remains the dominant OpenPGP implementation; age (`https://age-encryption.org`, Filippo Valsorda) is a modern minimal file-encryption tool with strong defaults; Signal (`https://signal.org`) is the reference deployment of the Signal Protocol; OpenMLS (`https://openmls.tech`) is the primary open MLS implementation. For reading lists beyond the core texts above, the Cryptography Engineering mailing list (`https://blog.cryptographyengineering.com`, Matthew Green) and Bruce Schneier's *Cryptogram* newsletter remain current and accessible.

[3]Further reading on cryptography as a discipline and its political history. For the foundational twentieth-century papers: Claude E. Shannon, "Communication Theory of Secrecy Systems," *Bell System Technical Journal* 28, no. 4 (1949): 656–715, `https://www.iacr.org/museum/shannon/shannon45.pdf`, established cryptography as a mathematical discipline and articulated the "enemy knows the system" principle (Kerckhoffs's principle, originally stated in Auguste Kerckhoffs, "La cryptographie militaire," *Journal des sciences militaires* 9 [January–February 1883]: 5–38); Whitfield Diffie and Martin Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* IT-22, no. 6 (1976): 644–654, invented public-key cryptography; Ronald Rivest, Adi Shamir, and Leonard Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM* 21, no. 2 (1978): 120–126, gave the first practical public-key scheme; David Chaum, "Security Without Identification: Transaction Systems to Make Big Brother Obsolete," *Communications of the ACM* 28, no. 10 (1985): 1030–1044, extended the discipline to privacy-preserving transactions. For the political and ethical frame, Phillip Rogaway, "The Moral Character of Cryptographic Work," IACR Distinguished Lecture (2015), `http://web.cs.ucdavis.edu/~rogaway/papers/moral-fn.pdf`, argues

that cryptographic design is an act of political engineering; Julian Assange, with Jacob Appelbaum, Andy Müller-Maguhn, and Jérémie Zimmermann, *Cypherpunks: Freedom and the Future of the Internet* (OR Books, 2012), collects the cypherpunk-political reading of the same discipline; Jon Callas's oft-quoted formulation that "cryptography is the ultimate form of non-violent direct action" captures the same position from a practitioner who helped ship PGP, iMessage, and Silent Circle. For history and narrative, David Kahn, *The Codebreakers: The Story of Secret Writing*, rev. ed. (Scribner, 1996; original 1967), is the definitive long-form history of cryptography through the mid-twentieth century; Simon Singh, *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography* (Anchor, 2000), covers similar ground accessibly. Steven Levy, *Crypto: How the Code Rebels Beat the Government, Saving Privacy in the Digital Age* (Viking, 2001), is the narrative history of public-key cryptography and the cypherpunk movement through PGP and the Clipper Chip fight.

[4]AES is specified in NIST FIPS 197, *Advanced Encryption Standard* (*AES*) (2001), `https://csrc.nist.gov/pubs/fips/197/final`. ChaCha20 is specified in RFC 8439, *ChaCha20 and Poly1305 for IETF Protocols* (2018), `https://www.rfc-editor.org/rfc/rfc8439`.

[5]Whitfield Diffie and Martin E. Hellman, "New Directions in Cryptography," *IEEE Transactions on Information Theory* 22, no. 6 (1976): 644–654.

[6]Ronald L. Rivest, Adi Shamir, and Leonard M. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM* 21, no. 2 (1978): 120–126.

[7]Victor S. Miller, "Use of Elliptic Curves in Cryptography," *Advances in Cryptology: CRYPTO '85* (1985): 417–426; Neal Koblitz, "Elliptic Curve Cryptosystems," *Mathematics of Computation* 48, no. 177 (1987): 203–209.

[8]The secp256k1 curve parameters are defined in Standards for Efficient Cryptography Group, *SEC 2: Recommended Elliptic Curve Domain Parameters*, version 2.0 (2010), `https://www.secg.org/sec2-v2.pdf`. The reference open-source implementation used by Bitcoin Core is libsecp256k1, `https://github.com/bitcoin-core/secp256k1`.

[9]National Institute of Standards and Technology, *Secure Hash Standard* (*SHS*), FIPS PUB 180-4 (Gaithersburg, MD: NIST, 2015).

[10]For digital signature foundations, see Shafi Goldwasser, Silvio Micali, and Ronald L. Rivest, "A Digital Signature Scheme Secure Against Adaptive Chosen-Message Attacks," *SIAM Journal on Computing* 17, no. 2 (1988): 281–308.

[11]Don Johnson, Alfred Menezes, and Scott Vanstone, "The Elliptic Curve Digital Signature Algorithm (ECDSA)," *International Journal of Information Security* 1, no. 1 (2001): 36–63. ECDSA is the signature scheme used in Bitcoin, Ethereum, and many TLS implementations; it applies elliptic curve point multiplication to produce signatures that are smaller and faster than RSA signatures at equivalent security levels. The original standard is ANSI X9.62 (1998); NIST's version is FIPS 186-5 (2023), `https://csrc.nist.gov/pubs/fips/186-5/final`.

[12]Claus-Peter Schnorr, "Efficient Signature Generation by Smart Cards," *Journal of Cryptology* 4, no. 3 (1991): 161–174. Schnorr signatures have a linear algebraic

structure that RSA and ECDSA lack: multiple signatures can be aggregated into a single signature whose size is independent of the number of signers (MuSig, MuSig2), and the scheme admits provably secure constructions in the random oracle model. Bitcoin added native Schnorr signatures via BIP 340 (activated in the Taproot upgrade, November 2021), `https://github.com/bitcoin/bips/blob/master/bip-0340.mediawiki`.

[13]The "$5 wrench attack" refers to the observation that physical coercion is often easier than cryptanalysis. See XKCD comic 538, "Security."

[14]Peter W. Shor, "Algorithms for Quantum Computation: Discrete Logarithms and Factoring," *Proceedings 35th Annual Symposium on Foundations of Computer Science* (1994): 124–134.

[15]Lov K. Grover, "A Fast Quantum Mechanical Algorithm for Database Search," *Proceedings of the 28th Annual ACM Symposium on Theory of Computing* (1996): 212–219. Grover's algorithm provides a quadratic speedup for unstructured search problems: a quantum computer can search an N-item database in $O(\sqrt{N})$ steps, compared to O(N) classically. Applied to symmetric cryptography, this halves the effective key length: AES-128 under Grover's attack has roughly the security of 64-bit classical search. The standard response is to double key lengths (AES-256 in place of AES-128), which restores the original security margin. Unlike Shor's algorithm, which breaks RSA and ECC asymptotically, Grover's attack leaves symmetric and hash-based primitives intact; it requires larger parameters but not architectural replacement.

[16]NIST finalized the core post-quantum standards in August 2024 (ML-KEM, ML-DSA, SLH-DSA) and added HQC as a code-based backup KEM in March 2025; Signal's PQXDH (Sept 2023), Apple's iMessage PQ3 (Feb 2024), Google Chrome's X25519MLKEM768 (Nov 2024), and Cloudflare's edge (30%+ PQ TLS by mid-2025) are the largest deployed hybrid implementations; NSA CNSA 2.0 and UK NCSC target 2033–2035 for high-value migration. NIST Post-Quantum Cryptography Standardization project, `https://csrc.nist.gov/projects/post-quantum-cryptography`. Finalized standards: FIPS 203 (ML-KEM), `https://csrc.nist.gov/pubs/fips/203/final`; FIPS 204 (ML-DSA), `https://csrc.nist.gov/pubs/fips/204/final`; FIPS 205 (SLH-DSA), `https://csrc.nist.gov/pubs/fips/205/final`. HQC selection: NIST IR 8547 (March 2025), `https://csrc.nist.gov/pubs/ir/8547/ipd`. Signal PQXDH specification at `https://signal.org/docs/specifications/pqxdh/`. Apple iMessage PQ3 documentation at `https://security.apple.com/blog/imessage-pq3/`. Google Chrome post-quantum TLS announcement at `https://blog.chromium.org/2024/05/advancing-our-amazing-bet-on-asymmetric.html` and X25519MLKEM768 rollout at `https://security.googleblog.com/`. Cloudflare post-quantum TLS telemetry at `https://blog.cloudflare.com/pq-2024/`. U.S. National Security Agency CNSA 2.0 suite, `https://media.defense.gov/2022/Sep/07/2003071834`

/-1/-1/0/CSA_CNSA_2.0_ALGORITHMS_.PDF. U.K. National Cyber Security Centre post-quantum migration guidance at https://www.ncsc.gov.uk/guidance/pqc-migration-timeline.

15

ZERO-KNOWLEDGE PROOFS

> "In cryptography, a zero-knowledge proof is a method by which one party can prove to another party that a given statement is true, without conveying any information apart from the fact that the statement is indeed true."
>
> Shafi Goldwasser, Silvio Micali, and Charles Rackoff, "The Knowledge Complexity of Interactive Proof Systems" (1985)[1]

Zero-knowledge proofs let one party prove that a statement is true without revealing any information beyond the truth of the statement itself. For privacy, this changes what verification requires. Traditional verification demands disclosure. Zero-knowledge verification avoids that demand.

This chapter examines how zero-knowledge proofs work and what they enable, along with their limitations. The technology remains complex, but its privacy implications are simple: verification without routine surveillance can become possible.

For a first pass, three comparison points are enough: proof size, trusted setup, and the practical cost of proving and verification. The later taxonomy separates constructions by how they answer those three questions.

15.1 The Verification Dilemma

Traditional Verification Requires Disclosure

Consider common verification scenarios. Identity checks expose identifying documents, so the verifier sees name, address, photo, document numbers, and more data than the verification requires. Age checks for alcohol purchases expose the exact birthdate and name when the merchant needs only confirmation of being over 21. Professional qualification checks expose issuing institution, dates, grades, and other credentials on the same document. Creditworthiness checks reveal income sources, spending patterns, account balances, and transaction history.

In each case, verification reveals more information than logically necessary for the verification's purpose.

The Privacy Cost

This over-disclosure has costs. Each verification adds to databases, and over time, entities accumulate complete profiles from individually minor disclosures. More data creates more targets; data breaches expose information that privacy-preserving verification would never have collected. Disclosed data enables inferences, and knowing someone's birthdate and zip code often suffices to uniquely identify them. Verifiers accumulate information while individuals cannot audit how it is used, creating systematic power imbalance.

The Fundamental Dilemma

Verification requires information, but providing information destroys privacy. Traditional systems force a choice between participating in verification-requiring activities or maintaining privacy.

Zero-knowledge proofs resolve this dilemma by separating what is proven from what is revealed.

15.2 The ZK Concept: Proving Without Revealing

Interactive Proofs

The original zero-knowledge concept involves interactive protocols.[1] A prover wants to convince a verifier of a statement without revealing underlying information. They engage in a protocol where the verifier poses challenges and the prover responds.

The tunnel analogy illustrates the concept. Picture a horseshoe-shaped tunnel with both ends opening onto the same clearing. Deep inside, where the two paths meet, sits a number-locked door. Alice claims she knows the combination. Bob stands in the clearing where he can see both tunnel exits but cannot see inside. Alice enters while Bob looks away. Bob then calls out "come out the left exit" or "come out the right exit." If Alice knows the combination, she can always comply, unlocking the door if necessary. If she does not, she is trapped on whichever side she entered and has only a 50% chance of guessing correctly. After twenty successful rounds, Bob is certain Alice knows the combination, yet he never learned what it is.

In that exchange, Bob learns only that Alice knows the secret and not the secret itself.

Non-Interactive Proofs

Modern applications typically use non-interactive zero-knowledge proofs (NIZKs). Instead of a back-and-forth protocol, the prover generates a single proof that anyone can verify.

Non-interactivity enables practical applications: proofs can be attached to transactions or stored on blockchains, and they can be verified by multiple parties without prover involvement.

Formal Properties

Zero-knowledge proof systems have three properties.[2] Completeness means that if the statement is true, an honest prover can convince an honest verifier. Soundness means that if the statement is false, no cheating prover can convince an honest verifier except with negligible probability. Zero-knowledge means the verifier learns

nothing beyond the truth of the statement; even after seeing the proof, the verifier cannot extract additional information. Soundness and zero-knowledge are in tension: strong soundness requires the proof to "contain" something about the statement; zero-knowledge requires the proof to reveal nothing. Cryptographic constructions achieve both properties through mathematical techniques that seem almost magical.

15.3 Types: SNARKs, STARKs, Bulletproofs

Different zero-knowledge proof systems make different tradeoffs. The main constructions in current use are:

SNARKs

Succinct Non-interactive Arguments of Knowledge (SNARKs) are proof systems whose proof size and verification cost grow sub-linearly, typically polylogarithmically, in the size of the computation being proved.[3] A SNARK can certify a computation of millions of steps with a proof of a few hundred bytes that verifies in milliseconds. This succinctness property is what makes blockchain rollups and shielded transactions possible: a single constant-size proof can stand in for an arbitrarily complex computation, so the verifier's work stays tiny no matter how large the proven computation gets. Practical Groth16 SNARKs produce proofs around 200 bytes; more recent constructions trade proof size for other properties including transparency and post-quantum security.

Traditional SNARKs require a "trusted setup": generating initial parameters through a ceremony that, if compromised, would allow fake proofs. Transparent SNARKs now exist that eliminate this requirement. Systems like Spartan[4] and Plonky2[5] achieve SNARK-like properties without trusted setup, using different cryptographic techniques. The tradeoff is typically somewhat larger proofs or slower proving times compared to trusted-setup SNARKs.

Most widely deployed SNARK systems rely on cryptographic assumptions vulnerable to quantum computing attacks. Zcash uses SNARKs for shielded transactions.[6]

STARKs

Scalable Transparent Arguments of Knowledge (STARKs) eliminate the trusted setup requirement.[7] STARKs offer transparency (no trusted setup), rely on minimal cryptographic assumptions (hash functions), are generally regarded as post-quantum, and scale well to large computations. The tradeoff: they produce larger proofs (tens to hundreds of kilobytes) and involve more complex verification.

StarkWare uses STARKs for Ethereum scaling solutions.

Bulletproofs

Bulletproofs are designed for range proofs and similar applications.[8] They require no trusted setup and produce moderate proof sizes, making them efficient for range proofs (proving a value is in a range without revealing it). Verification is slower than SNARKs, and their use is narrower than that of SNARKs or STARKs.

Monero uses Bulletproofs, and now Bulletproofs+, primarily for range proofs inside confidential transactions.

Choosing Among Constructions

The constructions differ across several dimensions. Traditional SNARKs produce tiny proofs (around 300 bytes) with fast verification but require a trusted setup and are not quantum-safe; transparent SNARKs eliminate the trusted setup at some cost to proof size or proving time. STARKs avoid the trusted setup and are generally regarded as post-quantum but generate much larger proofs (around 100 kilobytes) with moderate verification speed. Bulletproofs occupy a middle ground with medium-sized proofs (around 2 kilobytes) and no trusted setup, though verification is slower and they lack quantum resistance.

The right choice depends on application requirements. Blockchain base layers typically favor small proofs such as SNARKs, applications requiring post-quantum security lean toward STARKs, and confidential-transaction range proofs often use Bulletproofs.

15.4 Applications: Zcash, Rollups, Identity

Private Cryptocurrency: Zcash

Zcash implements shielded transactions using SNARKs.[9] Users can transact with both counterparties hidden and the amount concealed. The blockchain records that a valid transaction occurred without revealing its details.

The proof shows that input coins exist and are unspent, that the sender possesses the spending keys, and that input and output values balance without double-spending. It does this without revealing which coins are spent or who receives funds, and without exposing the amount transferred.

Zcash shows zero-knowledge proofs deployed at scale in adversarial conditions. Shielded adoption remains partial, and transparent or selectively transparent usage still limits privacy in practice. Many exchanges require transparent addresses for deposits and withdrawals, undermining the privacy benefits. The shielding technology works, but network effects and regulatory pressure limit its use.

Zcash also shows the risks of complex cryptography and the need for ongoing maintenance. In 2019, a vulnerability was discovered in the original Sprout proving system that could have allowed undetectable inflation. The bug existed since launch; no exploitation was detected, and the Sapling upgrade (deployed in 2018) had already fixed it before the vulnerability was publicly disclosed. The Zcash team's practice of deploying cryptographic upgrades allowed the fix to precede public knowledge of the bug. The discovery showed that even carefully designed systems can harbor critical bugs, making active protocol development essential. Zcash continues to undergo upgrades; the protocol is actively maintained and improved.

Blockchain Scalability: Validity Rollups

Zero-knowledge proofs enable "rollups" that scale blockchain throughput.[10] A rollup executes many transactions off-chain, then posts a proof to the main chain that all executions were valid.

The main chain verifies only the proof, not the individual transactions. This compresses many transactions into one proof while letting the main chain validate correctness and preserve data availability so transaction data can still be reconstructed.

Systems such as StarkNet and zkSync use this approach for Ethereum scaling.

Identity and Credentials

Zero-knowledge proofs enable privacy-preserving identity verification. The core insight is that most verification situations require proving a property, not revealing the underlying data: a bar needs to know you are over 21, not your exact birthdate, and a lender needs to know you can repay, not your complete financial history. A service needs to know you are a licensed professional, not your home address.

Traditional verification bundles necessary proof with unnecessary disclosure. Showing a driver's license to prove age reveals name, address, birthdate, driver's license number, and photo. Each of these data points creates risk: identity theft, targeted harassment, and database breach exposure. The verification accomplished its purpose; the disclosure exceeded its purpose.

Zero-knowledge credentials invert this relationship. Age verification allows proving that age exceeds a threshold without revealing birthdate; the proof shows "birthdate $<$ (today - 21 years)" without exposing the birthdate itself. In a well-designed system, the verifier learns what they need (legal age confirmed) and not the surrounding personal data. Credential verification allows proving possession of a valid credential (degree, license, membership) without revealing identifying details; the verifier learns the credential is valid without learning who holds it. An employer can verify professional licensure without learning the candidate's home address or when they obtained the license. Selective disclosure allows proving only specific properties from a rich credential containing many attributes; from a driver's license, one can prove only "state of residence" without revealing personal identifying details.

Several projects are implementing these concepts. Polygon ID (now Privado ID) enables issuance and verification of credentials where holders can generate zero-knowledge proofs showing credential properties.[11] A credential holder can prove group membership or age qualification, and can also show possession of a valid license, without revealing the credential itself or their identity. The system separates the credential, which contains rich data, from the proof, which reveals only what the verifier needs.

zkPass takes a different approach, enabling zero-knowledge proofs from existing Web2 documents.[12] Users can prove properties from driver's licenses and utility bills, as well as financial records, without uploading those documents to a third party. The proof generation happens on the user's device; the verifier receives only the proof, not the underlying document. This bridges the gap between existing identity infrastructure and zero-knowledge verification.

Current know-your-customer (KYC) requirements force users to surrender full identity documents to every service requiring verification. Each disclosure creates a database that can be breached or subpoenaed, then misused. Zero-knowledge KYC (ZK-KYC) offers an alternative: prove compliance without repeated downstream disclosure. A user could prove "identity verified by licensed institution" without revealing the identity itself to downstream services. The compliance requirement is met downstream without recreating full identity databases at each service.

This remains largely speculative as deployed regulation. Regulators accustomed to full disclosure may resist verification methods they cannot inspect. But the technical capability exists to satisfy regulatory purposes while minimizing privacy invasion. Whether regulators will accept ZK-KYC depends on policy choices, not technical limitations.

The privacy implications extend beyond individual transactions. Identity verification currently creates complete profiles as a side effect. Each age check and credential presentation, along with address verification, adds to dossiers that follow individuals across contexts. Zero-knowledge verification sharply limits this accumulation. Each

proof reveals only what that specific verification requires, leaving far less data to aggregate across contexts.

Not all projects in this space equally prioritize privacy. Worldcoin, now World,[13] uses zero-knowledge proofs to verify "unique human" status while still depending on biometric enrollment. Current project materials say the biometric pipeline is designed to avoid a centralized biometric database, but the system still depends on trust in the surrounding hardware, operators, and enrollment design. The technology enables privacy; whether implementations provide privacy depends on the full system design.

ZK over Government-Issued Documents

The largest deployment of zero-knowledge identity attestation in 2025 runs on existing government-issued travel documents. The ICAO 9303 biometric passport standard places a signed NFC chip in roughly 1.8 billion passports worldwide. The chip contains the holder's name, date of birth, nationality, and document number, each signed by the issuing country's certificate authority. The signature is ordinarily verified by border-control equipment. A zero-knowledge circuit over the same signature lets a mobile device prove specific attributes from the chip without revealing the chip's full contents or the holder's identity.

Self (self.xyz, formerly Proof of Passport) is the flagship production deployment, using on-device NFC scanning with a ZK circuit over the Document Security Object to produce proofs of age, nationality, or sanctions-list non-membership; the project reports over one hundred million verifications across its deployments, including integrations with Google and with blockchain-based applications that accept the attestation as a sybil-resistance primitive. Comparable constructions run on the Aztec L2 (zkPassport), on Russia's anti-Putin referendum infrastructure (Rarimo's Freedom Tool), and against India's Aadhaar records (Anon Aadhaar).[14]

The approach uses the state's existing identity infrastructure against its surveillance uses. The government signed the document to authenticate the holder at border control; the ZK circuit over the signature uses the authentication without the identification,

which is exactly the separation this chapter has been arguing zero-knowledge makes available. The countermove available to the state is to restructure document issuance: change signature schemes, shorten certificate lifetimes, or add attestation requirements the circuit cannot satisfy without revealing more than it currently does. The approach works today because hundreds of millions of existing documents are already in circulation with signatures the circuits can verify, and that installed base cannot be recalled quickly.

Privacy Pass and Anonymous Credentials at Scale

Privacy Pass is the IETF-standardized anonymous credential protocol (RFCs 9576 and 9577) that ships in iOS 16 and later as Apple Private Access Tokens.[15] The construction is a blind signature over an attestation: an identity provider (Apple) attests that a device is a real user and not a bot, and a verifier (Cloudflare or Fastly) accepts the attestation without learning which device the attestation covers. The blind signature prevents the verifier from linking across requests. The attestation covers device integrity at a single moment, not the user's identity, behavior, or persistent activity.

Apple Private Access Tokens are active on hundreds of millions of Apple devices and are accepted by Cloudflare in place of CAPTCHAs on a significant fraction of the web. No prior anonymous-credential deployment has approached this scale. The unlinkability the blind signature provides against the verifier is real, and the attestation the system depends on is issued by a single corporation whose participation is what makes the scale possible. Mass deployment of anonymous credentials arrived through a centralized attester, not around one.

ZK Privacy Pools on Smart-Contract Chains

Shielded pools on smart-contract chains have moved from research to production since 2023. Railgun operates across Ethereum, BSC, Polygon, and Arbitrum as an opt-in privacy pool using Groth16 SNARKs.[16] Aztec Network operates as a private-by-default L2 where the execution environment itself preserves privacy, with se-

lective disclosure controlled by each user. Privacy Pools, proposed by Buterin and co-authors in 2023, extend Tornado Cash's deposit-mixer architecture with association sets: a user proves membership in a subset of the pool's deposits that excludes deposits the user wishes to exclude on compliance grounds, and the proof reveals nothing about which deposit in the approved subset is the user's.[17] The 0xbow.io deployment is the production reference.

The three designs differ in how much privacy they default to and in how privacy interacts with compliance pressure. Railgun is an opt-in pool that users must choose to shield. Aztec is shielded by default. Privacy Pools are the design that most directly addresses the Tornado Cash criticism: privacy without opacity to a specific list of excluded deposits, with the exclusion list chosen by the user and not by any single authority. The construction shows that the zero-knowledge primitive can be used to distinguish criminal use from legitimate use without a central authority making the distinction, which is a property the older privacy-pool designs did not claim.

Current vs. Speculative Applications

Currently deployed and working applications include Zcash shielded transactions and rollup-based scaling of Ethereum and similar platforms. Zero-knowledge-based identity and attestation tooling has moved during the last two years from demonstration into production for several specific patterns. Group-membership attestation, which lets a participant prove membership in a registered group without revealing which member they are, has been deployed at conference-scale events for ticketing, voting, and access control; Semaphore and Zupass are the deployed examples of this pattern. OAuth-backed wallet creation has been deployed on at least one mainnet blockchain through zkLogin, which lets a user derive a wallet from an existing Google or Apple account login without the account provider learning which wallet was derived. Zero-knowledge selective disclosure over standard verifiable credentials has been implemented for narrow deployments including age-gate attestations and professional-license verification.[18]

Universal verifiable computation and zero-knowledge voting at scale, along with full privacy-preserving compliance systems, remain speculative in the book's sense that their practical constraints still rule out broad deployment. The pattern that has moved into production in every case is the narrow one, where a specific statement is attested under a specific trust model with specific performance. The pattern that remains research is the general-purpose one. The direction of movement is what matters for the book's framework: the category of production deployments is expanding, the performance envelope is widening, and the primitives the research community developed a decade ago are now carrying real load.

The technology is real and deployed in important applications, but maturity differs sharply by use case. A reader deciding whether to rely on a specific zero-knowledge system for a specific purpose should evaluate the maturity of that specific system, not the maturity of the field as a whole.

15.5 Economic Implications: Verification as Service

Once proof can be separated from disclosure, verification itself becomes a distinct service. Complex proofs require substantial computation to generate even though verification is fast, which opens room for specialized proving infrastructure. The privacy tradeoff is that a prover that sees the witness can reintroduce the trust problem the proof was meant to narrow; verification markets fit the book's framework only when witness handling stays local, is split across multiple parties, or is protected by hardware that keeps the prover from extracting the underlying data. Where those conditions hold, zero-knowledge systems can reduce verification costs across the economy: compliance without exposure of business details, verifiable credentials without fraud-checking overhead, aggregate statistics without routine dossier building. The gains are narrow today and directional over time, verification need not always mean surveillance.

15.6 Limitations

Trusted Setup Requirements

Many SNARK constructions require a "trusted setup": an initial ceremony generating parameters that all future proofs and verifications use. If the ceremony is compromised (someone retains the "toxic waste" used to generate parameters), they can create fake proofs that verify correctly.

Trusted setups come in several varieties with different tradeoffs. Per-circuit setups, as used in Zcash's original Sprout system, require a new ceremony for each application or circuit change; any modification to the computation being proved requires discarding the old parameters and running a new trusted setup. Universal setups, as in systems like PLONK[19] and Marlin,[20] generate parameters that work for any circuit up to a certain size; one ceremony can serve many applications, though the ceremony itself remains a trust assumption. Transparent setups eliminate the requirement entirely: STARKs and some newer SNARKs like Spartan derive their parameters from public randomness, requiring no ceremony and leaving no toxic waste that could be retained.

Mitigations for systems requiring trusted setup include multi-party computation ceremonies where only one participant needs to be honest and Powers of Tau ceremonies with thousands of participants. But applications with strict trust requirements may prefer transparent constructions despite their larger proofs.

The trusted setup is a real limitation that responsible implementations take seriously.

Computational Intensity

Proof generation is computationally expensive. Simple proofs take seconds; complex proofs can take minutes or hours. This limits applications where latency matters.

Verification is fast (often milliseconds), but generation is the bottleneck. Hardware acceleration (GPUs, FPGAs, ASICs) can help but does not eliminate the fundamental computational cost.

Implementation Complexity

Zero-knowledge proof systems are mathematically sophisticated, and implementation bugs can destroy the security properties the mathematics is supposed to guarantee. A soundness failure lets invalid proofs verify, which means forged statements pass as true; a zero-knowledge failure leaks information through the proof itself, which defeats the privacy the system exists to provide. Subtle implementation errors of both kinds have affected deployed systems, often undetected for years before being discovered.

Auditing is correspondingly hard. Few engineers can competently review ZK implementations at the level needed to catch these bugs, and the scarcity of qualified reviewers concentrates trust in a small community whose time is itself a bottleneck on how many systems can be meaningfully examined.

What ZK Proofs Cannot Solve

Chapter 14 examined what cryptography cannot solve in general: endpoint security, metadata exposure, physical coercion, key authenticity. Zero-knowledge proofs inherit these limitations and add ZK-specific constraints.

The most significant is the garbage-in-garbage-out problem, often called the oracle problem. A zero-knowledge proof verifies that a computation was performed correctly on given inputs. It says nothing about whether those inputs were true. An application that claims "this loan is collateralized" using a ZK proof has proven only that the computation was done correctly on the price data it received; if that data was stale or manipulated, or false, the proof is meaningless. No amount of cryptographic sophistication can bridge the gap between "correctly computed" and "true." Proving you processed oracle data correctly does not prove the oracle provided accurate data. This limitation is fundamental: ZK proofs verify computation, not reality.

Zero-knowledge adds complexity that may be unnecessary. When the verification problem does not require hiding information, simpler cryptographic tools suffice. Adding ZK machinery where it

is not needed increases attack surface and implementation difficulty without corresponding benefit.

Chapter Summary

Zero-knowledge proofs separate verification from disclosure. Completeness, soundness, and zero-knowledge together ensure that true statements can be proven, false statements cannot, and the proof reveals nothing beyond the truth of the statement, the verifier learns what they need without access to the underlying data. The primitive resolves the verification dilemma that otherwise forces a choice between participation and privacy. Different constructions make different tradeoffs. SNARKs produce tiny proofs with fast verification but traditionally require trusted setup and are vulnerable to quantum attack; transparent SNARKs eliminate the setup at some cost to proof size or proving time. STARKs also avoid trusted setup and provide quantum resistance but produce much larger proofs. Bulletproofs occupy a middle ground suited to range proofs. Application requirements determine the appropriate choice.

Production deployment has reached specific patterns. Zcash shielded transactions and validity rollups for blockchain scaling are operational. Group-membership attestation through Semaphore and Zupass runs at conference-scale events, Sui's zkLogin derives wallets from OAuth logins without the provider learning the derivation, and W3C Verifiable Credentials support selective disclosure in narrow age-gate and professional-licensing deployments. Zero-knowledge circuits over ICAO 9303 passport signatures, through Self and related projects, carry hundreds of millions of verifications using identity infrastructure the state itself issued. Privacy Pass ships as Apple Private Access Tokens on hundreds of millions of iOS devices. Railgun, Aztec, and Privacy Pools carry shielded transactions on smart-contract chains, with association-set constructions addressing the compliance concerns that older pool designs did not.

Universal verifiable computation and zero-knowledge KYC at broad economic scale remain speculative. The oracle problem constrains every application that depends on external data: a ZK

proof verifies that computation was performed correctly on given inputs, not that the inputs were true. The category of production deployments is expanding, the performance envelope is widening, and the primitives developed in the research community a decade ago now carry real load, but maturity differs sharply by use case, and a reader evaluating any specific system should evaluate that system, not the field as a whole. The communication-layer and monetary-layer privacy that ZK proofs complement belong to Chapters 17 and 20 respectively.[21]

Notes

[1]Shafi Goldwasser, Silvio Micali, and Charles Rackoff, "The Knowledge Complexity of Interactive Proof Systems," *SIAM Journal on Computing* 18, no. 1 (1989): 186–208.

[2]For accessible introduction to zero-knowledge proofs, see Matthew Green, "Zero Knowledge Proofs: An Illustrated Primer," *A Few Thoughts on Cryptographic Engineering* (2014).

[3]Jens Groth, "On the Size of Pairing-Based Non-interactive Arguments," *Advances in Cryptology: EUROCRYPT 2016* (2016): 305–326.

[4]Srinath Setty, "Spartan: Efficient and General-Purpose zkSNARKs without Trusted Setup," *Advances in Cryptology: CRYPTO 2020* (2020): 704–737, `https://eprint.iacr.org/2019/550`. Spartan uses the sum-check protocol over multilinear extensions to achieve transparent SNARKs with no trusted setup and no toxic waste; security rests on the discrete-logarithm assumption in a group of prime order. The proving complexity is O(n log n) for arithmetic circuits of size n. Spartan is the foundation for Microsoft's Nova folding-scheme prover (`https://github.com/microsoft/Nova`) and influenced subsequent transparent SNARK designs.

[5]Polygon Zero Team, "Plonky2: Fast Recursive Arguments with PLONK and FRI" (2022), `https://github.com/0xPolygonZero/plonky2/blob/main/plonky2/plonky2.pdf`. Plonky2 combines PLONK's universal-setup arithmetization with FRI (Fast Reed-Solomon Interactive Oracle Proofs of Proximity) as the polynomial commitment scheme, eliminating the elliptic-curve pairings that conventional SNARKs require. The result is a transparent SNARK with recursive proof composition times under 170 milliseconds on consumer hardware; Polygon uses it as the proving backend for its zkEVM and other scaling infrastructure. FRI's hash-function-only assumptions make Plonky2 plausibly post-quantum, though formal post-quantum security proofs for the full construction remain an open research question.

[6]Zcash, `https://z.cash`, is a privacy-preserving cryptocurrency launched in October 2016 and derived from the Zerocash protocol described in Eli

Ben-Sasson et al., "Zerocash: Decentralized Anonymous Payments from Bitcoin," *IEEE Symposium on Security and Privacy* (2014), https://eprint.iacr.org/2014/349. Zcash shielded transactions use zk-SNARKs (originally BCTV14, later Groth16, now Halo 2 since the 2022 NU5 upgrade which removed the trusted setup) to prove transaction validity without revealing sender, receiver, or amount. The protocol is developed by the Electric Coin Company (https://z.cash) with the Zcash Foundation (https://zfnd.org) as an independent public-interest counterpart. For the broader design space of privacy-preserving cryptocurrencies, Monero (https://www.getmonero.org) takes a different approach: ring signatures and stealth addresses provide its anonymity guarantees; see Shen Noether et al., "Ring Confidential Transactions," *Ledger* 1 (2016), https://ledgerjournal.org/ojs/ledger/article/view/34. Mimblewimble (implemented in Grin and Beam) and CoinJoin-based approaches on Bitcoin (Wasabi, JoinMarket) are the other main families.

[7]Eli Ben-Sasson et al., "Scalable, Transparent, and Post-Quantum Secure Computational Integrity," *CRYPTO 2018* (2018).

[8]Benedikt Bünz et al., "Bulletproofs: Short Proofs for Confidential Transactions and More," *IEEE Symposium on Security and Privacy* (2018): 315–334.

[9]Zerocash protocol: see note 4 above for the full citation and Zcash deployment details.

[10]For rollup designs, see Vitalik Buterin, "An Incomplete Guide to Rollups," *Ethereum Blog* (2021). The two largest ZK-rollup deployments on Ethereum are StarkNet (https://www.starknet.io/, developed by StarkWare) and zkSync (https://zksync.io/, developed by Matter Labs); both use validity proofs to compress large batches of transactions into a single on-chain proof. StarkNet documentation at https://www.starknet.io/; zkSync documentation at https://zksync.io/. Both use validity proofs to compress large numbers of transactions into a single on-chain proof.

[11]Polygon ID (rebranded as Privado ID) documentation at https://docs.privado.id/. See also "Introducing Polygon ID, Zero-Knowledge Identity for Web3," Polygon Technology Blog (2022), and Identity Foundation, "Guest Blog: Polygon ID: Verifiable Credentials Meet Web3."

[12]zkPass documentation at https://docs.zkpass.org/. zkPass uses Three-Party TLS (3P-TLS) and Multi-Party Computation to enable zero-knowledge proofs from HTTPS website data without exposing raw documents.

[13]Worldcoin (rebranded as World, https://world.org/) uses iris-scan biometric enrollment to issue World ID, a zero-knowledge proof of unique human status. Sam Altman founded the project; Tools for Humanity is the operating entity. The privacy model depends on trust in the Orb hardware and the enrollment operator; technical architecture at https://whitepaper.worldcoin.org/. For critical analysis, see the data-protection investigations by Spanish (AEPD) and Kenyan regulators in 2023-2024.

[14]Self (formerly Proof of Passport), zkPassport, Rarimo, and Anon Aadhaar use zero-knowledge circuits over ICAO 9303 passport NFC signatures and Aadhaar RSA signatures to produce selective-disclosure identity attestations from government-issued credentials; Rarimo's Freedom Tool ran a production deployment for a Russian-opposition referendum in 2024. Self Protocol (formerly Proof of Passport), `https://self.xyz/`, with source at `https://github.com/selfxyz`. zkPassport, `https://zkpassport.id/`. Rarimo and Freedom Tool, `https://rarimo.com/` and `https://freedomtool.org/`. Anon Aadhaar, a Privacy and Scaling Explorations project, `https://pse.dev/projects/anon-aadhaar`, with source at `https://github.com/anon-aadhaar`. ICAO Doc 9303 Machine Readable Travel Documents specification, `https://www.icao.int/publications/Documents/9303_p11_cons_en.pdf`. For the broader category of ZK attestation over existing signatures, see the academic line starting from Jens Groth, Rafail Ostrovsky, and Amit Sahai, "Perfect Non-Interactive Zero Knowledge for NP," *EUROCRYPT 2006*, and continuing through Noir, Circom, and the proof-system developer toolchains cited in note 16.

[15]Privacy Pass (IETF RFCs 9576 and 9577) is the anonymous-credential standard, and Apple Private Access Tokens are the single largest production deployment; the architecture's scale depends on Apple acting as centralized attester. RFC 9576 (Privacy Pass Architectural Framework), Alex Davidson, Jana Iyengar, Christopher A. Wood, `https://datatracker.ietf.org/doc/rfc9576/`. RFC 9577 (Privacy Pass Issuance Protocols), Steven Valdez, `https://datatracker.ietf.org/doc/rfc9577/`. Apple Private Access Tokens documentation at `https://developer.apple.com/documentation/authenticationservices/privacy_access_tokens`, with Cloudflare integration documented at `https://blog.cloudflare.com/eliminating-captchas-on-iphones-and-macs-using-new-standard/`. The earlier Cloudflare Privacy Pass research at `https://blog.cloudflare.com/privacy-pass-v3/` and Alex Davidson, Ian Goldberg, Nick Sullivan, George Tankersley, and Filippo Valsorda, "Privacy Pass: Bypassing Internet Challenges Anonymously," *PoPETs* 2018, no. 3 (2018).

[16]Railgun documentation at `https://railgun.org/`, with the protocol specification at `https://docs.railgun.org/`. Aztec Network documentation at `https://aztec.network/`, with the Noir privacy-first programming language at `https://noir-lang.org/`. For the comparison with Zcash's shielded-pool architecture examined in note 4, see Aztec's "Private Execution Environment" documentation at `https://docs.aztec.network/aztec/concepts/`.

[17]Vitalik Buterin, Ameen Soleimani, Jacob Illum, Matthias Nadler, Fabian Schär, and Ameen Soleimani, "Blockchain Privacy and Regulatory Compliance: Towards a Practical Equilibrium" (2023), `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4563364`, introduces the association-set construction. The 0xbow deployment is at

https://0xbow.io/, with source at https://github.com/0xbow-io. For the Tornado Cash predecessor that Privacy Pools is a response to, see *Van Loon v. Department of the Treasury*, 122 F.4th 549 (5th Cir. 2024), and Chapter 13 for the legal-phase analysis the construction intersects with.

[18]Semaphore (group-membership proofs) and Zupass (event-scale ticketing/credentials) are the deployed examples of zk-based group attestation; Sui's zkLogin lets users derive wallets from OAuth login without the OAuth provider learning the derivation; verifiable-credential selective disclosure is in narrow production for age-gate and professional-license use cases. Semaphore, https://semaphore.pse.dev/, with source at https://github.com/semaphore-protocol/semaphore. Zupass (PCD framework), https://pcd.dev/ and https://zupass.org/. Mina PCD / Proof-Carrying Data documentation at https://pcd.dev/docs. zkLogin on Sui, https://docs.sui.io/concepts/cryptography/zklogin, with specification at https://sui.io/zklogin. World Wide Web Consortium Verifiable Credentials Data Model, https://www.w3.org/TR/vc-data-model/. ISO/IEC 18013-5 mobile driver's license standard, https://www.iso.org/standard/69084.html. On age-gate and professional-license attestation deployments, see Privado ID, https://docs.privado.id/, and the eIDAS 2.0 European Digital Identity Wallet reference implementation at https://github.com/eu-digital-identity-wallet.

[19]Ariel Gabizon, Zachary J. Williamson, and Oana Ciobanu, "PLONK: Permutations over Lagrange-Bases for Oecumenical Noninteractive Arguments of Knowledge," *IACR ePrint* 2019/953 (2019), https://eprint.iacr.org/2019/953. PLONK introduced a universal and updateable structured reference string: a single trusted setup ceremony produces parameters usable by any arithmetic circuit up to a given size, and the parameters can be updated by additional participants without rerunning the ceremony from scratch. This made universal-setup SNARKs practical for ecosystems like Ethereum, where many independent applications share proving infrastructure; PLONK arithmetization is now the basis for Halo 2, UltraPLONK, and numerous zkEVM backends.

[20]Alothman Zac, Mary Maller, and Pratyush Mishra, "Marlin: Preprocessing zkSNARKs with Universal and Updatable SRS," *Advances in Cryptology: EUROCRYPT 2020* (2020): 738–768, https://eprint.iacr.org/2019/1047. Marlin generalizes the universal-setup approach of PLONK using a different arithmetization (R1CS over a universal SRS based on polynomial commitments) and provides the first formal proof of knowledge-soundness for a universal preprocessing zkSNARK. The construction underlies the Aleo blockchain's Leo programming language and the snarkVM execution environment, which use Marlin as their proving backend for general-purpose private computation.

[21]Further reading on zero-knowledge proofs. The foundational papers are Shafi Goldwasser, Silvio Micali, and Charles Rackoff, "The Knowledge Complexity of Interactive Proof-Systems," *SIAM Journal on Computing* 18, no. 1 (1989): 186–208

(originally STOC 1985), https://people.csail.mit.edu/silvio/Selected%20Scientific%20Papers/Proof%20Systems/The_Knowledge_Complexity_Of_Interactive_Proof_Systems.pdf; and Oded Goldreich, Silvio Micali, and Avi Wigderson, "Proofs That Yield Nothing But Their Validity" (1991), which extended ZK to all of NP. For the textbook treatment, Oded Goldreich, *Foundations of Cryptography: Volume 1, Basic Tools* (Cambridge University Press, 2001), chapter 4, and *Volume 2, Basic Applications* (2004), chapter 4, are the canonical references. Jonathan Katz and Yehuda Lindell, *Introduction to Modern Cryptography*, 3rd ed. (CRC Press, 2020), covers ZK at the graduate-course level. For the practitioner side, Justin Thaler, *Proofs, Arguments, and Zero-Knowledge* (Now Publishers, 2023), https://people.cs.georgetown.edu/jthaler/ProofsArgsAndZK.pdf, is the modern reference on SNARK/STARK construction. Matthew Green's "Zero Knowledge Proofs: An Illustrated Primer" at https://blog.cryptographyengineering.com/2014/11/27/zero-knowledge-proofs-illustrated-primer/ is the best accessible explanation. For SNARKs specifically, the zkSNARK series by Vitalik Buterin at https://vitalik.eth.limo/general/2016/12/10/qap.html is practitioner-friendly. The 0xPARC curriculum at https://0xparc.org is a community resource for learning ZK programming hands-on. On the practical tooling, Circom (https://docs.circom.io), Halo 2 (https://zcash.github.io/halo2/), arkworks-rs (https://github.com/arkworks-rs), and Nova (https://github.com/microsoft/Nova) are the reference implementations for SNARK development; for STARKs, StarkWare's Cairo (https://www.cairo-lang.org) is the primary public toolkit. On the identity-and-credentials side, the W3C Verifiable Credentials Data Model (https://www.w3.org/TR/vc-data-model/) and Hyperledger AnonCreds (https://github.com/hyperledger/anoncreds-spec) define the standards that much ZK identity work targets.

16
COMPUTING ON SECRETS

"The problem of privacy is not how to hide data,
but how to use data without disclosing it."

Andrew C. Yao, "Protocols for Secure Computations" (1982)[1]

The previous chapter examined zero-knowledge proofs, which let one party prove a statement is true without revealing why. This chapter is about the adjacent and complementary capability: performing a computation on secret inputs, without any participant in the computation learning those inputs.

A zero-knowledge proof operates on a result that already exists. The computation whose correctness is proved was performed somewhere, by someone, on inputs that were available somewhere. The proof certifies what was computed, not who saw the inputs. Zero-knowledge is the verification primitive.

Computing on secrets is the execution primitive. It answers a different question: can a computation be performed when no single party is permitted to see the inputs? The answer is yes, under three architectural families that have reached practical deployment within the last decade. Fully homomorphic encryption performs arithmetic directly on ciphertexts. Secure multi-party computation splits a computation across participants so that no participant sees the whole input. Trusted execution environments run computation inside a hardware enclave whose contents are opaque to the operating system and to the hardware owner.

These families are complementary to zero-knowledge proofs, not competing with them. A pipeline that protects privacy end-to-end typically uses computing-on-secrets to operate on private data and zero-knowledge to prove the result was produced honestly. Chapter 15 examined the proving half of that pipeline. This chapter examines the executing half.

16.1 The Verification-Execution Distinction

Consider a voting system in which each voter's choice must remain private, and in which the tallied result must be publicly verifiable. The inputs (individual ballots) must remain secret; the output (the tally) must be computable and verifiable.

Plain encryption handles the secrecy of inputs at rest but not their use. If ballots are encrypted and sent to a counting service, the counting service must decrypt them to count them, which means at least one party has access to every individual ballot. The architecture has not solved the problem; it has only moved the problem from transport to a specific processing node.

Zero-knowledge proofs handle the verifiability of the output without directly addressing the privacy of the inputs. A zk-proof can certify that a tally was computed correctly from some set of ballots. The proof alone does not ensure that the ballots were private during the tally; it ensures only that the tally corresponds to them. If the counting service decrypted ballots to produce the tally, the proof is compatible with that architecture. Privacy of inputs is a separate requirement.

Computing on secrets handles the privacy of inputs directly. Homomorphic addition lets the counting service add up ballots in ciphertext form without ever decrypting any individual ballot; the final tally ciphertext is decrypted once by a distributed threshold of key holders. Secure multi-party computation distributes the counting across nodes so that no single node sees any voter's choice, and the output is reconstructed only from the joint result. A trusted execution environment runs the count inside an enclave that refuses to disclose its memory to any external party and provides an attesta-

tion that the canonical counting software ran on the expected ballot inputs.

The complete system combines the two families. Ballots remain private through computing-on-secrets; and the tally is publicly verifiable through zero-knowledge. Neither family alone solves the problem; their combination does.

16.2 Homomorphic Encryption: Arithmetic on Ciphertext

Homomorphic encryption permits operations on ciphertexts that correspond to operations on the underlying plaintexts. If Enc(a) is the encryption of a plaintext value a, and Enc(b) is the encryption of b, then a homomorphic addition produces a ciphertext whose decryption equals a + b, without any party learning either value in plaintext form. The construction is almost miraculous when stated plainly, and it is achievable under specific mathematical assumptions.

From Partial to Fully Homomorphic

The earliest homomorphic schemes supported only a single operation. RSA famously supports homomorphic multiplication, which is a property its designers did not advertise but which researchers quickly noticed and which motivated the broader line of research. The Paillier cryptosystem supports homomorphic addition. Partially homomorphic schemes have been used in practice for narrow applications for decades, including e-voting systems and private information retrieval.[2]

Craig Gentry's 2009 dissertation was the first construction of a fully homomorphic encryption scheme capable of both addition and multiplication on arbitrary circuits.[3] The original construction was roughly a billion times slower than the equivalent plaintext computation, which ruled it out for any serious application. The theoretical breakthrough was that such a scheme existed at all. The engineering program that followed has reduced the overhead by roughly four

orders of magnitude over fifteen years, which moves narrow applications into the range of feasibility while keeping general-purpose computation out of reach.

The Current Deployment Field

The present constructions that practitioners use fall into three families. The Brakerski-Gentry-Vaikuntanathan (BGV) and Brakerski-Fan-Vercauteren (BFV) schemes support integer arithmetic and are well-suited to computations that can be expressed as polynomials over finite fields.[4] The Cheon-Kim-Kim-Song (CKKS) scheme supports approximate arithmetic over the real numbers, which is well-suited to machine-learning inference where small numerical errors are acceptable. The Torus FHE (TFHE) family supports fast evaluation of arbitrary Boolean circuits through a technique called programmable bootstrapping, which is well-suited to logic-heavy computations. Open-source libraries including OpenFHE, Microsoft SEAL, and Zama's TFHE-rs implement these schemes in production-quality code.[5]

The practical deployments fall into categories whose common property is that narrow usefulness beats broad slowness. Encrypted machine-learning inference lets a client send an encrypted input to a service and receive an encrypted output without the service seeing the input; CKKS is the usual vehicle, and the applications include encrypted medical diagnosis and encrypted financial-risk scoring. Private information retrieval lets a client query a database without revealing which record it wants; variations of homomorphic techniques power deployments of this kind in privacy-focused email and search products. Encrypted databases let a service host encrypted customer data and answer queries over it without decrypting; BGV and BFV are the usual vehicles.

What homomorphic encryption does not yet do is support fully general-purpose computation at acceptable speed. The overhead remains large enough that applications are chosen for their tolerance for latency and their limited computational complexity. The frontier is moving toward broader applicability, and improved hardware

acceleration has contributed to the move, but the architecture is not yet the default way to run arbitrary programs on private data.

The Structural Promise

Beneath any specific application sits a simpler claim. Homomorphic encryption breaks the assumption that the service running a computation must see the data on which the computation runs. That assumption had been treated as a law of cloud architecture for two decades. Homomorphic encryption denies it. The service and the user become parties to an exchange in which the service performs a function and the user retains epistemic ownership of the inputs. The contract structure that was previously infeasible is now infeasible only at certain workload sizes, and the workloads for which it is feasible are expanding each year.

16.3 Secure Multi-Party Computation

Secure multi-party computation (MPC) addresses a different problem with a different architecture. Multiple parties each hold private inputs; they wish to jointly compute a function of those inputs and to learn only the output. No party learns any other party's input beyond what the output itself reveals.

The Core Primitives

The classical constructions rely on a small family of cryptographic primitives. Secret sharing, originally studied by Shamir in the 1970s, splits a secret into shares such that any threshold number of shares reconstructs it but any smaller subset reveals nothing. A function to be computed on secret inputs can be expressed as an arithmetic circuit, and operations on the circuit can be performed on shares, with each participant computing local operations on their share and jointly opening only the final output. Garbled circuits, introduced by Yao, allow two parties to evaluate a Boolean circuit where one party encrypts the circuit and the other evaluates the encryption; each gate's behavior is correct but its intermediate val-

ues are hidden. Oblivious transfer lets a party receive one of two values chosen by another party without the sender learning which value was chosen; it is a foundational primitive on which many MPC protocols rest.

The practical constructions combine these primitives in ways that fit specific applications. The GMW, BGW, and SPDZ protocol families each offer different tradeoffs among efficiency, security against actively malicious adversaries, and the threshold of honest participants required. A production MPC deployment must pick a protocol whose threat model matches the deployment's actual threat model, and the choice determines how many participants can collude without compromising the computation.

Threshold Signing in Production

The single most successful MPC deployment to date is threshold signing, and it has reached institutional scale. A threshold signature scheme distributes the private key across multiple parties such that signing requires a threshold of them to cooperate, and no subset below the threshold can produce a valid signature. No party ever holds the full private key, which means a breach of any individual party does not compromise the system. Fireblocks, the largest institutional custody provider, operates threshold-signed wallets for roughly two thousand customers including major banks and payment processors; the architecture is the basis on which several of those institutions extended credit into the crypto-asset space at all, because the alternative of a single-party custodian required levels of counterparty trust they were not prepared to extend.

Two Schnorr-based constructions have made threshold signing directly viable on Bitcoin. MuSig2, specified in BIP 327, aggregates an arbitrary number of signers into a single public key and a single signature indistinguishable from a normal single-party Schnorr signature. FROST (Flexible Round-Optimized Schnorr Threshold signatures) extends the same approach to t-of-n threshold signing, so any t of the n key shareholders suffice to produce a signature and the protocol tolerates up to n-t shares going offline or hostile without losing liveness. Both became usable on Bitcoin with the Taproot

activation in November 2021; MuSig2 and FROST produce valid BIP 340 Schnorr signatures, and the resulting on-chain transactions are indistinguishable from single-signer spends. Chain analysis cannot distinguish a ten-of-ten MuSig2 transaction from a single-key transaction, nor a three-of-five FROST spend from either. The architecture delivers MPC's distributed-trust guarantee and Bitcoin's base-layer privacy in one construction, which is the combination the book's framework treats as the target for every privacy-preserving primitive.

Threshold signing is also the basis for several consumer-oriented wallet products that offer the user security properties stronger than single-key custody without the usability burden of multi-signature setups at the base-layer protocol. The protocols that enable these products rely on the same MPC primitives as the institutional products, adapted for the specific constraints of mobile devices and retail-user onboarding.[6]

The Structural Promise

MPC distributes trust across parties who do not trust one another, and the distributed trust produces a cryptographic guarantee that no single party can violate. This is a different guarantee from the one homomorphic encryption provides. Homomorphic encryption lets one party execute a computation on another party's private input. Multi-party computation lets several parties execute a computation on their own private inputs, with no party learning any other party's input. The two primitives solve complementary problems, and many real systems combine them.

Parties with private information can now cooperate on joint tasks that benefit from that information without pooling it in a central repository or extending trust to a shared intermediary. The cryptographic guarantee replaces a coordination cost that was previously priced into every such cooperation: the cost of establishing enough trust to share the inputs. When that cost falls, cooperation expands to cases that were previously uneconomic.

16.4 Trusted Execution Environments

Trusted execution environments (TEEs) take a different architectural path. Where cryptographic means prevent the computing party from seeing the inputs, a TEE prevents the computing party from seeing the inputs through hardware isolation instead. A region of memory inside the processor is marked as an enclave; code running inside the enclave can decrypt inputs, operate on them, and produce outputs, while code running outside the enclave (including the operating system, the hypervisor, and any other process on the machine) cannot observe the enclave's memory. An attestation mechanism allows a remote party to verify that a specific piece of software is running inside an authentic enclave on an authentic processor, giving the remote party confidence that the computation was performed as described.

The Platforms

Intel Software Guard Extensions (SGX) was the first mainstream TEE and remains the most widely deployed in server applications. SGX was deprecated on Intel's consumer processor line starting with the eleventh-generation Core products, in recognition of a series of side-channel attacks that made client-side SGX impractical. Its server variant, now called Scalable SGX, remains available on Xeon processors and continues to support cloud-scale applications. Intel Trust Domain Extensions (TDX) is the successor architecture, designed to address the side-channel vulnerabilities through a combination of hardware and software changes. AMD offers Secure Encrypted Virtualization with Secure Nested Paging (SEV-SNP), which provides enclave-like properties at the virtual-machine boundary. ARM TrustZone has been the primary mobile-device TEE for over a decade and underlies the biometric and payment-credential storage on hundreds of millions of devices. Apple's Secure Enclave is a hardened variant of ARM TrustZone that also handles device-local cryptographic operations.[7]

The Consumer-Scale Deployment

TEE architecture at consumer scale breaks into two deployment patterns. The first is the device-local enclave. ARM TrustZone on Android phones, Apple's Secure Enclave on iOS devices, and the TPM 2.0 specification on modern PCs together place a hardened execution environment inside hundreds of millions of consumer devices, where it holds biometric templates, payment-credential keys, disk-encryption keys, and attestation material. The device-local enclave is the oldest and most widely deployed consumer TEE pattern, and it is the least controversial because the user and the enclave are colocated in the same physical device under the same owner.

The second pattern is the remote attested enclave. A consumer device sends a request to a cloud-operated server whose hardware and software can be cryptographically attested and whose code is published for inspection. Each server's software is cryptographically measured, the measurement is published, and the device refuses to send a request to a server whose measurement does not match a published image. The servers are built to hold no persistent state beyond the request's lifetime, so that even a full compromise of a server at a later time yields no earlier request's inputs. The enterprise cloud tier has offered this architecture for several years through Google Cloud Confidential Computing, Microsoft Azure Confidential Computing, and AWS Nitro Enclaves, which expose attested-enclave primitives to business workloads built on AMD SEV-SNP, Intel TDX, and similar hardware. Signal's Private Contact Discovery uses an SGX-based enclave to let Signal clients look up which of their phone contacts use Signal without revealing the query to the Signal server. Apple Private Cloud Compute, introduced in June 2024 to serve workloads from Apple Intelligence that exceed on-device compute, is the most consumer-visible instance to date and is the deployment whose architecture has been published in the most detail. The category has been converging across vendors for several years; the Apple release made it legible to a mass consumer audience.

The architecture is worth describing in detail because it illustrates both the promise and the limit of TEE-based privacy. The promise is that a large-scale cloud service can be operated in a way that the service itself cannot observe user inputs, which is the same structural promise as homomorphic encryption but achieved through different means. Apple's design specifically targets the case where Apple itself is the adversary: an append-only transparency log records the cryptographic measurement of every production server image, the user's device refuses to encrypt requests to any node whose attestation does not match a logged measurement, and the signed binaries are published for independent researchers to reproduce. The residual trust is narrow but real: the silicon vendor's attestation root keys, the manufacturing supply chain that installs them, and the presence of independent monitors watching the transparency log for split views or suppressed entries. Side-channel attacks against the enclave's hardware isolation remain a separate risk class. Trust is reduced, not eliminated, and the user has traded trust in a provider's operational practices for trust in its hardware root and in the auditors watching its log.

Attested Inference: Frontier Models Without Disclosure

The remote-attested-enclave pattern has found a second application whose privacy stakes are higher than most of its early uses. Large language models have concentrated in the hands of a few operators because frontier-scale inference requires compute that individual users do not have. A prompt sent to a hosted model is a revealed preference handed to the operator in plaintext, and the default cloud architecture gives the operator full view of every request every customer makes. The surveillance consequences compound across professions (medical queries, legal drafts, source-protected journalism, negotiation strategy, personal correspondence) in a way base-layer communication surveillance does not, because a model prompt is often a condensed statement of what its author is thinking at that moment.

Two architectures have moved attested inference from proposal into production. Nvidia introduced confidential-computing sup-

port on H100 GPUs in 2023, reached general availability in April 2024, and extended the capability across H200 and Blackwell B200 with multi-GPU Protected PCIe in April 2025.[8] The GPU establishes a hardware root of trust with a device-unique key burned in at manufacture, boots a measured firmware image, and carries on an attested SPDM session with a CPU TEE (AMD SEV-SNP or Intel TDX) that relays encrypted data in and out of the GPU. A relying party verifies the attestation chain through Nvidia's remote attestation service and sends prompts whose plaintext the GPU alone sees. Azure made this generally available for customers in September 2024 as NCC H100 v5 confidential VMs; Google Cloud shipped the equivalent on A3 Confidential instances. A set of third-party operators run consumer- and developer-facing inference services on this hardware. Tinfoil publishes device attestations on every request and binds model identity to the attestation; Edgeless Systems' Privatemode serves EU-hosted confidential inference on the same Azure substrate; Nillion's nilAI deploys open-weight models (Llama 3, DeepSeek, Gemma, GPT-OSS) inside attested GPUs. The open-weight frontier (Llama 3.3 70B, DeepSeek R1, Kimi K2, GPT-OSS 120B) now runs end-to-end inside attested enclaves at production scale.

The deployment that goes furthest integrates confidentiality with anonymity. Apple Private Cloud Compute combines the attested-enclave architecture described above with two orthogonal primitives lifted from anonymous-communication research. Every production request routes through an OHTTP relay (RFC 9458) operated by a third party that strips the client IP before the request reaches Apple's infrastructure, so the cloud side never sees the network origin of the request. Authorization uses RSA blind signatures (RFC 9474), a single-use credential that proves the request comes from a legitimate Apple device without identifying which device, so the cloud side never sees the user identity either. The combination means Apple's production side cannot read the prompt (confidentiality, from the TEE), cannot link two prompts to the same user (unlinkability, from the fresh blind credential per request), and cannot identify the user at all (anonymity, from the OHTTP strip).

Target-diffusion routing prevents the load balancer from steering a specific user's traffic to a compromised node, because it has no user identifier to key on. The sovereign-host tier examined in Chapter 23 provides the same three guarantees by never leaving the user's hardware at all; PCC provides them at the cloud layer for workloads sovereign-host hardware cannot run.

The gap sits at the closed frontier. OpenAI, Anthropic, and Google ship contractual zero-retention tiers and tenant-isolated deployments, not hardware-attested inference. Microsoft's published collaboration with OpenAI on confidential Whisper inference (September 2024) is a validation exercise, not a general-availability product for the flagship models. The structural reason is that closed model weights become a policed asset once exposed to a customer-facing TEE boundary, and the operators have so far preferred contractual guarantees over architectural ones. Users who require an architectural guarantee today have two options: run open-weight inference inside an attested enclave (Tinfoil, Privatemode, or the same cloud primitives), or accept the lower-capability alternative of running the model locally on their own hardware. The closed frontier will either move to attested inference as customer demand accumulates or concede that segment of the market to the open-weight stack that already has.

TEE-based privacy is the primitive that scales the most easily today. Homomorphic encryption remains expensive for general computation. Multi-party computation requires at least two non-colluding parties, and organizing them is a social problem as much as a technical one. TEEs run at plaintext speeds on general-purpose workloads, and the cryptographic overhead is limited to attestation and to the isolation boundary. The primitive fits the cloud-service business model almost exactly.

The qualifier is that the trust reduction is not a trust elimination. Every TEE architecture depends on the integrity of the hardware, the attestation mechanism, and the software image against which attestation is verified. The history of side-channel attacks against SGX shows that hardware integrity is contested, not given; the recent examples of Spectre, Meltdown, and their variants show that the

contested boundary extends across the entire processor design.[9] The user who evaluates a TEE-based privacy claim must evaluate the specific hardware vendor's track record, the specific attestation architecture's design, and the specific software image published for verification. None of these evaluations is trivial, and none of them can be reduced to a single binary. TEEs are useful when the alternative is no privacy guarantee at all, and they are inferior when the alternative is a cryptographic primitive that does not require trusting hardware.

16.5 Private Information Retrieval

Private Information Retrieval (PIR) lets a client retrieve an item from a server's database without the server learning which item was retrieved. The query index is the private input; the item is the output. The server performs a computation over its entire database that depends cryptographically on the query but preserves no observable trace of which element the query selected.

The primitive comes in two families. Information-theoretic PIR replicates the database across multiple non-colluding servers; the client splits the query into shares and each server answers its share without learning the full query. Information-theoretic privacy holds as long as at least one server remains honest. Computational PIR operates with a single server and relies on cryptographic assumptions, typically additively homomorphic encryption or fully homomorphic encryption: the client sends an encrypted query vector, the server computes over its database in ciphertext, and the response ciphertext decrypts to the requested item without the server learning which index produced it. The cost is that every query must touch every record in the database, which is why computational PIR took two decades of engineering to become practical.

Practical deployments have arrived over the last five years. SealPIR, from Microsoft Research and Stanford in 2018, showed computational PIR over gigabyte-scale databases at single-query latencies of hundreds of milliseconds. SimplePIR and DoublePIR, introduced in 2022, cut the per-query cost by an order of magnitude

through offline preprocessing, reaching throughput in the tens of gigabits per second and making PIR viable for high-volume services. Spiral, OnionPIR, and FrodoPIR occupy adjacent points in the design space. Signal's private contact discovery combines PIR-style queries with a trusted execution environment so the server has cryptographic assurance it cannot see which phone numbers a client is looking up.

The applications that matter for the book's framework are the ones in which the query is itself the information. Domain-name resolution is the canonical case: every DNS lookup tells the resolver which services a user is about to contact, which aggregates into an exhaustive record of online activity. Oblivious DNS over HTTPS (ODoH) mitigates this by splitting the query across an HTTP proxy and a DNS resolver, but a full PIR-based resolver would hide the query from both. Bitcoin light clients are the most directly relevant case for the book's argument. BIP 37 bloom filters leak which addresses a wallet cares about; BIP 158 compact block filters leak less but are more expensive in regards to bandwidth and computation. A PIR-based light client would let the wallet fetch transaction data without telling the serving node which addresses, UTXOs, or transactions it is watching. Lightning light-client flows reach Lightning Service Providers that observe every payment activity the client performs; PIR-aware routing queries would materially reduce that exposure. Certificate-revocation checks through OCSP reveal which TLS connections a browser is about to establish, and OCSP stapling partially masks this without eliminating it; a PIR-based revocation service would close the gap.

The structural promise is narrower than homomorphic encryption or MPC but sharply focused. PIR does not compute; it retrieves. What it adds to the book's stack is the ability to interact with remote databases in a way that does not leak the query. Every service the privacy architecture depends on is also a lookup surface, and every lookup surface is an observation channel unless the queries are themselves protected. PIR is the primitive that closes the query channel.[10]

16.6 Differential Privacy and the Aggregation Layer

Differential privacy is a statistical guarantee about aggregated data releases. Unlike the preceding primitives, it does not protect individual records during a computation; it protects the privacy of individuals whose records contribute to a released aggregate.

The Motivating Failure

The naive approach to sharing a dataset is to remove obvious identifiers (names, addresses, phone numbers) and publish what remains. This approach has failed repeatedly. In 1997, Latanya Sweeney re-identified the medical record of the Massachusetts governor from a "de-identified" state employee dataset by cross-referencing ZIP code, date of birth, and sex with publicly available voter rolls. In 2008, Arvind Narayanan and Vitaly Shmatikov re-identified a substantial fraction of users in the Netflix Prize dataset of 100 million movie ratings by cross-referencing with public IMDb reviews. The general pattern is the linkage attack: combine the released dataset with any auxiliary information the attacker has access to, and the combined information often suffices to re-identify individuals. No amount of column removal prevents this. The data itself encodes enough structure that uniqueness emerges even from coarse attributes.[11]

Differential privacy is the statistical response to the linkage attack. Instead of trying to remove identifying information from records, it changes what the release itself depends on. If the released aggregate would look the same whether any particular individual's record was included or not, then the released aggregate cannot be used to learn anything specific about that individual. The guarantee is a statement about the mechanism that produces the output, not about the output itself.

The Randomized-Response Intuition

Randomized response, introduced by Stanley Warner in 1965 for survey research, is the simplest case and makes the mechanism concrete. Suppose a researcher wants to know what fraction of

respondents have engaged in some sensitive activity that no one will admit to openly. The researcher asks each respondent to flip a coin privately. On heads, the respondent answers the sensitive question truthfully. On tails, the respondent flips again: on heads they answer "yes" regardless, on tails they answer "no" regardless. Each respondent's answer gives the researcher some information, but nothing definitive: a "yes" could have come from the truthful-yes path or from the random-yes path, and only the respondent knows which. Plausible deniability is built into the protocol.

Yet in aggregate the researcher can recover the true rate. If the true fraction of yeses is p, the expected fraction of observed yeses is $(1/2)p + 1/4$: on the truthful half the real rate comes through, and on the random half a coin flip gives "yes" half the time. Solving for p gives $p = 2(\text{observed fraction} - 1/4)$. The noise cancels in expectation; the individual deniability remains.

Randomized response is the prototype for every differentially private mechanism. Randomness at the individual level provides deniability; structure at the population level preserves utility. Real deployments generalize this tradeoff to arbitrary queries through a single privacy budget that bounds how much any one record can shift a released output and that composes additively across queries, so each release consumes part of the budget.

Central and Local Differential Privacy

Two architectures exist for where the noise is added. In central differential privacy, participants send their raw records to a trusted aggregator, the aggregator computes the true answer, and the aggregator adds noise before releasing the result. The user trusts the aggregator to handle the raw data correctly; the released aggregate is differentially private relative to external observers. The U.S. Census Bureau's TopDown Algorithm is a central-DP deployment: the Bureau receives raw census records, computes tabulations, and publishes noisy tabulations under a total privacy budget allocated across the many tables that make up the decennial release.

In local differential privacy, participants add noise to their own records before sending them anywhere. No trusted aggregator is

required. The cost is that much more noise must be added, because each record's own noise is the only noise protecting that record; the per-record noise must be large enough that a single record reveals almost nothing, even though noise averages out in the aggregate. Apple's telemetry system uses local DP: iOS devices perturb individual usage records on-device before transmitting them, and Apple aggregates the perturbed records without ever seeing the raw ones. The tradeoff is stark. Local DP provides stronger privacy (no trusted aggregator) at substantial accuracy cost; central DP provides better accuracy at the cost of requiring a trusted aggregator.

What Differential Privacy Protects and Does Not Protect

DP protects against two broad categories of attack. The first is membership inference: an adversary trying to determine whether a specific individual is in the dataset. Under a properly calibrated privacy budget, the adversary's advantage over guessing is bounded and shrinks as the budget tightens. The second is attribute inference about any named individual: learning anything specific about one person that the adversary did not already know. The same budget bounds how much the adversary can learn there as well.

DP does not protect against population-level inferences that do not depend on any one individual. If the data establishes that everyone in a demographic group buys a particular product, DP-released statistics will still reveal that pattern, and any individual known to be in that group will be implicated by the population-level inference. This is the "secrecy of the sample" limitation: DP protects the contribution of each individual to the inference, not the inference itself. DP also does not protect raw data at rest. If the raw dataset is held somewhere in plain form, a breach of that storage is not mitigated by the fact that a DP-released tabulation is protected.

Deployments

The U.S. Census Bureau applied central differential privacy to the 2020 census redistricting data release through the TopDown Algorithm, allocating a total ε budget across the tables and publish-

ing the noisy tabulations alongside the algorithm documentation so that statisticians can model the noise. Apple has used local differential privacy for aggregate telemetry (emoji usage, QuickType predictions, Safari crash reports) since iOS 10, with published per-type ε budgets. Google has used differential privacy for Chrome browser usage statistics, for the COVID-19 Community Mobility Reports,[12] and for audience measurement in its advertising products, deploying central, local, and hybrid variants depending on the product.

Where It Fits

Differential privacy belongs in the book's framework as the protection that applies when the population as a whole is the object of analysis and any individual's record contributes only as a data point. It complements the primitives earlier in the chapter, because it does not protect the individual computation on a sensitive input. Many real systems need both guarantees at once: that individual computations do not leak their inputs, and that aggregate releases do not deanonymize the individuals behind them. The combination is the architecture of a privacy-preserving analytics system, and the analytics system is the category in which every mature deployment of these primitives must settle.

16.7 The Design Space and the Complements

Each primitive fits a distinct region of the design space. Homomorphic encryption suits the case where one party runs a workload on another's private inputs and the overhead is tolerable. Secure multi-party computation suits the case where several parties each hold private inputs and can be organized around a specified collusion model. A trusted execution environment suits the general-purpose workload where cryptographic overhead is unacceptable and the hardware trust model is acceptable. Private information retrieval suits remote database access in which the query index is itself the private input. Differential privacy is the guarantee that

applies to population-level analysis; the contribution of any single record stays mathematically bounded.

The combinations matter as much as the individual primitives. Homomorphic encryption inside a trusted execution environment gives the hardware as a second defense against a broken scheme. Multi-party computation with participants running inside their own TEEs reduces the collusion surface to hardware. PIR over a TEE-hosted database combines query privacy with server-side integrity, which is the architecture Signal's private contact discovery uses. Zero-knowledge proofs attached to any of these primitives certify correctness without adding another execution copy. The frontier is the composition work that fits these primitives together for specific applications.

16.8 The Axiom of Resistance Applied to Compute

The book's Axiom of Resistance holds that a system's security is measured by the cost required to compromise it. Computing on secrets extends the measurement to the compute layer of the stack.

Homomorphic encryption's resistance surface is the set of cryptographic assumptions underlying the specific scheme in use. For lattice-based schemes, the assumption is learning-with-errors;[13] for code-based schemes, it is the hardness of decoding random linear codes; for other constructions, other assumptions apply. The assumption set is standard in modern cryptography, and it is the same assumption set that underlies the post-quantum standards in Chapter 14. A break in a specific scheme would affect the deployed applications using that scheme but would not break the family, because alternative schemes under different assumptions exist.

Multi-party computation's resistance surface is the threshold model. The guarantee holds if and only if fewer than the threshold number of participants collude. The threshold must be chosen large enough that collusion is infeasible, and the participants must be chosen to have diverse incentives. The guarantee is a function of the social structure of the participant set, not only of the cryptographic construction.

Trusted execution environments have the largest resistance surface. Hardware vendor integrity, attestation mechanism correctness, side-channel resistance, and software-image verification are each separate assumptions, and the composed guarantee is the conjunction of all of them. A break in any single assumption compromises the whole. The empirical track record of TEEs includes multiple documented side-channel attacks, and the architectural response has been to rebuild the hardware where software patching was insufficient. The history shows both that the primitive is contestable and that the contesting process is active.

Differential privacy's resistance surface is the privacy budget and the composition of multiple releases over time. Each release consumes a portion of the budget, and sufficient releases deanonymize the population regardless of how tightly any individual release was set. The practical deployments track and limit the cumulative budget, which requires operational discipline beyond the initial algorithmic choice.

The axiom's logic applies to each family independently and to compositions. An attacker must compromise each primitive in the composition, and the compromise cost compounds. The defender's architectural choice is the selection of primitives whose compositions raise the compromise cost above the adversary's willingness to pay. Computing on secrets does not promise invulnerability. It promises that the cost curve of observation now bends upward at the execution layer, which is the layer the adversary had previously assumed was free.

Chapter Summary

Computing on secrets is the execution complement to zero-knowledge verification. Where zero-knowledge proves correctness of a computation whose inputs are private, computing-on-secrets performs computation on inputs that the executing party cannot see. The two primitives solve complementary problems, and real privacy-preserving systems combine them. Three architectural families cover the practical design space. Fully homomorphic encryption

performs arithmetic on ciphertexts, breaking the assumption that a service running a computation must see the data the computation runs on; its overhead has fallen by four orders of magnitude since Gentry's 2009 breakthrough and continues to fall, though general-purpose computation remains out of reach. Secure multi-party computation distributes a computation across participants such that no party sees the whole input, with threshold signing as the most deployed production case, MuSig2 and FROST produce Schnorr signatures indistinguishable from single-sig spends on Bitcoin, and Fireblocks-class institutional custody runs threshold signing at scales that would otherwise require counterparty trust regulated institutions were not prepared to extend. Trusted execution environments run computation inside hardware enclaves whose contents are opaque to the hardware owner, with ARM TrustZone and Apple Secure Enclave on hundreds of millions of consumer devices and Apple Private Cloud Compute, Azure Confidential Computing, AWS Nitro Enclaves, and Google Cloud Confidential Computing carrying cloud-scale services.

Private information retrieval covers remote database access: every lookup surface is an observation channel unless queries are themselves protected, and PIR is the primitive that closes it. Signal's production deployment and the SealPIR and SimplePIR class of single-server implementations have made PIR practical where the query is itself the information. Differential privacy covers the aggregation layer, bounding through a single privacy budget how much an adversary can learn about any individual from a released statistic. Randomized response provides the intuition, calibrated noise provides the deployed construction, and the U.S. Census Bureau's TopDown Algorithm and Apple's local-DP telemetry are the largest deployed cases.

None of these primitives provides absolute guarantees. Homomorphic encryption relies on cryptographic assumptions that could fail, multi-party computation relies on threshold assumptions about collusion, trusted execution environments rely on hardware vendors whose integrity is contested, and differential privacy relies on budget discipline that can be exceeded. Maturity differs

sharply by use case: homomorphic encryption's overhead rules out general-purpose computation, multi-party computation's organizational requirements limit its deployment, and trusted execution environments have a history of side-channel vulnerabilities that constrains which threat models they suit. The Axiom of Resistance extends to this layer. Each primitive has a specific resistance surface, and compositions compound the compromise cost. Computing on secrets does not promise invulnerability; it bends the observation cost curve upward at the execution layer, and that bending is the architectural basis on which the privacy-preserving applications of the next decade will be built. Implementation requires expertise in the specific scheme, protocol, or platform, which the reading recommendations point to.

Notes

[1]Yao's 1982 paper founded secure computation; the epigraph paraphrases its orienting principle, not its exact language. Andrew Chi-Chih Yao, "Protocols for Secure Computations," *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science* (1982): 160–164.

[2]RSA's homomorphic multiplication property is implicit in the original paper: R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-Key Cryptosystems," *Communications of the ACM* 21, no. 2 (1978): 120–126. The Paillier cryptosystem's additive homomorphism is the subject of Pascal Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," *EUROCRYPT 1999*, Lecture Notes in Computer Science vol. 1592, pp. 223–238.

[3]Gentry's 2009 dissertation proved fully homomorphic encryption was possible; fifteen years of engineering have cut its overhead by four orders of magnitude and still counting. Craig Gentry, "Fully Homomorphic Encryption Using Ideal Lattices," *Proceedings of the 41st Annual ACM Symposium on Theory of Computing* (2009): 169–178. Gentry's Stanford dissertation (2009) is the longer treatment. For the broader survey of the field as it stands, see Craig Gentry, "Computing Arbitrary Functions of Encrypted Data," *Communications of the ACM* 53, no. 3 (2010): 97–105, and Zvika Brakerski, "Fundamentals of Fully Homomorphic Encryption," in *Providing Sound Foundations for Cryptography* (ACM Books, 2019), which remains the best accessible overview.

[4]The BGV scheme's originating paper, which established the framework for leveled fully homomorphic encryption over integers using ring learning-with-errors: Zvika Brakerski, Craig Gentry, and Vinod Vaikuntanathan, "Fully Homomorphic Encryption without Bootstrapping," *ITCS 2012*,

https://eprint.iacr.org/2011/277. The BFV variant that optimizes for integer arithmetic in practice: Junfeng Fan and Frederik Vercauteren, "Somewhat Practical Fully Homomorphic Encryption," *IACR ePrint* 2012/144, https://eprint.iacr.org/2012/144. These two papers are the theoretical foundation for the BGV/BFV library implementations cited in endnote 4; the libraries (Microsoft SEAL, OpenFHE) implement these schemes but the originating papers define the security reductions and parameter choices that determine practical security levels.

[5]BGV, BFV, CKKS, and TFHE are the four scheme families deployed in practice; OpenFHE, Microsoft SEAL, and Zama are the production-quality open-source implementations. OpenFHE, https://www.openfhe.org/. Microsoft SEAL, https://github.com/microsoft/SEAL. Zama (TFHE-rs, Concrete), https://www.zama.ai/, with source at https://github.com/zama-ai. On the CKKS scheme, Jung Hee Cheon, Andrey Kim, Miran Kim, and Yongsoo Song, "Homomorphic Encryption for Arithmetic of Approximate Numbers," *ASIACRYPT 2017*, is the originating paper. On TFHE and programmable bootstrapping, Ilaria Chillotti, Nicolas Gama, Mariya Georgieva, and Malika Izabachène, "TFHE: Fast Fully Homomorphic Encryption over the Torus," *Journal of Cryptology* 33, no. 1 (2020): 34–91.

[6]Secret sharing (Shamir 1979) and garbled circuits (Yao 1986) are the foundational MPC primitives; Fireblocks runs threshold-signing MPC for roughly two thousand institutional custody customers; on Bitcoin, MuSig2 (BIP 327) aggregates n-of-n Schnorr signers into one signature and FROST extends the same construction to t-of-n, with both producing on-chain signatures indistinguishable from single-party spends. Adi Shamir, "How to Share a Secret," *Communications of the ACM* 22, no. 11 (1979): 612–613, on secret sharing. Andrew C. Yao, "How to Generate and Exchange Secrets," *Proceedings of the 27th Annual Symposium on Foundations of Computer Science* (1986): 162–167, on garbled circuits. Oded Goldreich, *Foundations of Cryptography*, Volume 2 (Cambridge, 2004), is the authoritative theoretical treatment, and David Evans, Vladimir Kolesnikov, and Mike Rosulek, *A Pragmatic Introduction to Secure Multi-Party Computation* (NOW Publishers, 2018), is the most accessible practitioner reference. Fireblocks, https://www.fireblocks.com/. On threshold signatures for Bitcoin and Ethereum, see Rosario Gennaro and Steven Goldfeder, "Fast Multiparty Threshold ECDSA with Fast Trustless Setup," *ACM CCS 2018*, which underlies much of the production tooling. MuSig2: Jonas Nick, Tim Ruffing, and Yannick Seurin, "MuSig2: Simple Two-Round Schnorr Multi-Signatures," *CRYPTO 2021*, https://eprint.iacr.org/2020/1261; specification at BIP 327, https://github.com/bitcoin/bips/blob/master/bip-0327.mediawiki; reference implementation in libsecp256k1, https://github.com/bitcoin-core/secp256k1. FROST: Chelsea Komlo and Ian Goldberg, "FROST: Flexible Round-Optimized Schnorr Threshold Signatures," *SAC 2020*, https://eprint.iacr.org/2020/852; IETF draft specification at https://datatracker.ietf.org/doc/draft-irtf-cfrg-frost/;

production implementations at https://github.com/ZcashFoundation/frost (Zcash Foundation) and https://github.com/coinbase/kryptology (Coinbase). Taproot/Schnorr underlying activation: BIP 340, "Schnorr Signatures for secp256k1," https://github.com/bitcoin/bips/blob/master/bip-0340.mediawiki, and BIP 341 Taproot, https://github.com/bitcoin/bips/blob/master/bip-0341.mediawiki.

[7]Intel SGX and TDX, AMD SEV-SNP, ARM TrustZone, and Apple's Secure Enclave are the deployed TEE platforms; Apple Private Cloud Compute (June 2024) is the clearest consumer-scale TEE architecture published in detail. Victor Costan and Srinivas Devadas, "Intel SGX Explained," IACR ePrint Archive 2016/086, is the canonical technical reference for SGX. Intel TDX documentation at https://www.intel.com/content/www/us/en/developer/tools/trust-domain-extensions/overview.html. AMD SEV-SNP white paper, David Kaplan, Jeremy Powell, and Tom Woller, "AMD Memory Encryption" (2021), https://www.amd.com/system/files/TechDocs/SEV-SNP-strengthening-vm-isolation-with-integrity-protection-and-more.pdf. ARM TrustZone overview at https://www.arm.com/technologies/trustzone-for-cortex-a. Apple Secure Enclave documentation at https://support.apple.com/guide/security/secure-enclave-sec59b0b31ff/web. Apple Private Cloud Compute architecture at https://security.apple.com/blog/private-cloud-compute/. Confidential Computing Consortium, https://confidentialcomputing.io/.

[8]Nvidia confidential computing on Hopper and Blackwell GPUs for attested LLM inference. Gavin Uberti and Steven Hecht, "Confidential Computing on NVIDIA H100 GPUs for Secure and Trustworthy AI," Nvidia Developer Blog (August 3, 2023), https://developer.nvidia.com/blog/confidential-computing-on-h100-gpus-for-secure-and-trustworthy-ai/, is the technical introduction. General availability on H100 announced in "Announcing Confidential Computing General Access on NVIDIA H100 GPUs" (April 25, 2024), https://developer.nvidia.com/blog/announcing-confidential-computing-general-access-on-nvidia-h100-tensor-core-gpus/. Multi-GPU Protected PCIe general availability announced in "Announcing NVIDIA Secure AI General Availability" (April 23, 2025), https://developer.nvidia.com/blog/announcing-nvidia-secure-ai-general-availability/, covering HGX H100 and H200 eight-GPU configurations required for frontier-scale workloads. Azure NCC H100 v5 confidential VMs reached general availability in September 2024; see Microsoft Azure Confidential Computing blog, "General Availability: Azure Confidential VMs with NVIDIA H100 Tensor Core GPUs" (September 24, 2024), https://techcommunity.microsoft.com/blog/azureconfidentialcomputingblog/general-availability-azure-confidential-vms-wit

h-nvidia-h100-tensor-core-gpus/4242644. Third-party attested-inference platforms include Tinfoil (https://tinfoil.sh, with technology documentation at https://tinfoil.sh/technology), Edgeless Systems Privatemode (https://www.privatemode.ai, built on the Continuum AI framework at https://www.edgeless.systems/products/continuum), and Nillion nilAI (https://nillion.com); each publishes attestation reports and serves open-weight models on Nvidia CC infrastructure. For the Apple Private Cloud Compute anonymity mechanisms referenced in the prose, the two relevant RFCs are Eric Kinnear et al., "Oblivious HTTP," RFC 9458 (January 2024), https://www.rfc-editor.org/rfc/rfc9458.html, and Frank Denis, Jonathan Hoyland, David Schinazi, and Christopher A. Wood, "RSA Blind Signatures," RFC 9474 (October 2023), https://www.rfc-editor.org/rfc/rfc9474.html. Apple's architecture ties them together in "Private Cloud Compute: A New Frontier for AI Privacy in the Cloud" (June 10, 2024), https://security.apple.com/blog/private-cloud-compute/. For the ongoing gap at closed frontier models, see Microsoft's September 2024 announcement of confidential Whisper inference with OpenAI at the Azure blog above, and OpenAI's enterprise Zero Data Retention documentation at https://platform.openai.com/docs/guides/your-data. As of early 2026, no flagship closed-frontier model (GPT-5, Claude 4.x, Gemini Ultra) ships a publicly announced generally available hardware-attested confidential inference product for customers.

[9]Spectre and Meltdown are the landmark 2018 CPU speculative-execution vulnerabilities that revealed the contested nature of processor-level isolation boundaries; their relevance to TEE security is that the same hardware features (branch prediction, out-of-order execution, cache hierarchy) that enable these attacks are present inside the TEE boundary and can be exploited from outside it. Paul Kocher et al., "Spectre Attacks: Exploiting Speculative Execution," *IEEE Symposium on Security and Privacy* (2019), https://spectreattack.com/spectre.pdf. Moritz Lipp et al., "Meltdown: Reading Kernel Memory from User Space," *USENIX Security Symposium* (2018), https://meltdownattack.com/meltdown.pdf. The dedicated SGX side-channel attack literature is extensive; representative papers include Jo Van Bulck et al., "Foreshadow: Extracting the Keys to the Intel SGX Kingdom with Transient Out-of-Order Execution," *USENIX Security* (2018), https://foreshadowattack.eu/, and Daniel Moghimi et al., "Downfall: Exploiting Speculative Data Gathering," *USENIX Security* (2023), https://downfall.page/. Intel's deprecation of client SGX and the architectural pivot to TDX is documented in Intel's product change notifications starting with 11th-generation Core processors.

[10]Private Information Retrieval lets a client fetch an item from a server's database without the server learning which item; information-theoretic PIR needs multiple non-colluding servers, computational PIR runs on one server at the cost

of cryptographic assumptions; SealPIR (2018), SimplePIR and DoublePIR (2022), and Signal's private contact discovery are the reference practical systems. Foundational PIR papers: Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan, "Private Information Retrieval," *Journal of the ACM* 45, no. 6 (1998): 965–981, on information-theoretic PIR; Eyal Kushilevitz and Rafail Ostrovsky, "Replication Is Not Needed: Single Database, Computationally-Private Information Retrieval," *FOCS 1997*, on computational PIR. SealPIR: Sebastian Angel, Hao Chen, Kim Laine, and Srinath Setty, "PIR with Compressed Queries and Amortized Query Processing," *IEEE Symposium on Security and Privacy 2018*, `https://eprint.iacr.org/2017/1142`, with implementation at `https://github.com/microsoft/SealPIR`. SimplePIR and DoublePIR: Alexandra Henzinger, Matthew M. Hong, Henry Corrigan-Gibbs, Sarah Meiklejohn, and Vinod Vaikuntanathan, "One Server for the Price of Two: Simple and Fast Single-Server Private Information Retrieval," *USENIX Security 2023*, `https://eprint.iacr.org/2022/949`. Spiral: Samir Jordan Menon and David J. Wu, "Spiral: Fast, High-Rate Single-Server PIR via FHE Composition," *IEEE Symposium on Security and Privacy 2022*. Signal's private contact discovery: Signal blog, "Technology preview: Private contact discovery for Signal," `https://signal.org/blog/private-contact-discovery/`, and the subsequent SGX-and-oblivious-operations architecture described at `https://signal.org/blog/signal-private-group-system/`. On Oblivious DNS over HTTPS (ODoH) as a weaker-than-PIR mitigation, see IETF RFC 9230, `https://www.rfc-editor.org/rfc/rfc9230.html`. On Bitcoin light-client query privacy, BIP 37 Bloom filter privacy analysis in Arthur Gervais et al., "On the Privacy Provisions of Bloom Filters in Lightweight Bitcoin Clients," *ACSAC 2014*; BIP 158 compact block filter specification at `https://github.com/bitcoin/bips/blob/master/bip-0158.mediawiki`; ongoing privacy analysis at `https://bitcoinops.org/`. On PIR-based light clients as a research direction, see the Ethereum Portal Network discussions and the "Vitalik on Ethereum light client privacy" thread at `https://ethresear.ch/`.

[11]Differential privacy guarantees aggregate releases cannot be used to deanonymize individual records within an ε budget; Warner's 1965 randomized response is the intuition, Dwork's mid-2000s formalization is the mathematical framework, and the 2020 U.S. Census, Apple telemetry (since iOS 10), and Google usage statistics are the largest deployed cases. Foundational formalization: Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," *TCC 2006*, `https://iacr.org/archive/tcc2006/38760266/38760266.pdf`. Authoritative reference: Cynthia Dwork and Aaron Roth, *The Algorithmic Foundations of Differential Privacy*, *Foundations and Trends in Theoretical Computer Science* 9, nos. 3–4 (2014): 211–407, `https://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf`. Randomized response origin: Stanley L. Warner, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias," *Journal of the American*

Statistical Association 60, no. 309 (1965): 63–69. On the motivating failures: Latanya Sweeney, "Simple Demographics Often Identify People Uniquely," Carnegie Mellon University, Data Privacy Working Paper 3 (2000), https://dataprivacylab.org/projects/identifiability/, and Arvind Narayanan and Vitaly Shmatikov, "Robust De-anonymization of Large Sparse Datasets," *IEEE Symposium on Security and Privacy 2008*, https://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf, the Netflix Prize de-anonymization paper. U.S. Census Bureau TopDown Algorithm documentation at https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/disclosure-avoidance.html. Apple's deployment documented in Differential Privacy Team, "Learning with Privacy at Scale," *Apple Machine Learning Journal* 1, no. 8 (2017), https://machinelearning.apple.com/research/learning-with-privacy-at-scale. Google's OpenDP library at https://github.com/google/differential-privacy. On theoretical tradeoffs: Salil Vadhan, "The Complexity of Differential Privacy," in *Tutorials on the Foundations of Cryptography* (Springer, 2017). For the "secrecy of the sample" limitation and the distinction between individual-level and population-level inference, Cynthia Dwork, "Differential Privacy: A Survey of Results," *TAMC 2008*, is the standard reference.

[12]Google's COVID-19 Community Mobility Reports used aggregated, differentially private location data drawn from Google Maps users to produce anonymized population-level mobility summaries during the pandemic; the methodology document describes the DP parameters applied. Google LLC, "COVID-19 Community Mobility Reports: Understand the Data," methodology documentation available at https://support.google.com/covid19-mobility/answer/9825414. Archived report data at https://www.google.com/covid19/mobility/. On Google's broader differential privacy infrastructure, see Miguel Guevara, "Helping public health officials combat COVID-19," Google Blog (April 3, 2020), and the open-source DP library at https://github.com/google/differential-privacy, which implements the mechanisms used across Chrome, Mobility Reports, and advertising measurement products.

[13]The learning-with-errors (LWE) problem, introduced by Oded Regev in 2005, is the hardness assumption underlying BGV, BFV, CKKS, and the NIST post-quantum standards for key encapsulation and digital signatures; its ring variant (RLWE) is the assumption underlying most deployed FHE. Oded Regev, "On Lattices, Learning with Errors, Random Linear Codes, and Cryptography," *STOC 2005*, https://cims.nyu.edu/~regev/papers/qcrypto.pdf; journal version in *Journal of the ACM* 56, no. 6 (2009). Vadim Lyubashevsky, Chris Peikert, and Oded Regev, "On Ideal Lattices and Learning with Errors over Rings," *EUROCRYPT 2010*, https://eprint.iacr.org/2012/230, introduced the ring variant that all practical FHE schemes use. The NIST post-quantum

standardization process selected LWE-based schemes (CRYSTALS-Kyber, CRYSTALS-Dilithium) as primary standards; see NIST IR 8413 (2022), https://nvlpubs.nist.gov/nistpubs/ir/2022/NIST.IR.8413.pdf.

17

ANONYMOUS COMMUNICATION NETWORKS

"We kill people based on metadata."
Michael Hayden, panel discussion, Johns Hopkins University (2014)[1]

Chapter 14 established cryptographic foundations. Encryption protects the content of communications. But encryption alone is insufficient for privacy.

The problem is metadata: information about communications, not their content. Who communicates with whom, when, how often, and for how long reveals patterns that surveillance can exploit. Even with perfect content encryption, metadata enables full monitoring of social networks, political associations, and personal relationships.

This chapter examines architectural solutions to the metadata problem at two layers. At the transport layer, VPNs, Tor and I2P, and mixnets each offer a different point on the latency-anonymity trade-off. At the application layer, a sequence of encrypted-messaging protocols from PGP through Signal to MLS has progressively strengthened what end-to-end encryption protects and what it still leaks. Neither layer solves the problem alone; the two compose.

17.1 The Problem: How the Internet Leaks Privacy

IP Addresses: Built-In Identifiers

The Internet Protocol was designed for reliability, not privacy. Every packet contains the sender's IP address in plaintext, visible to every router along the path. The exposure is not a bug but a fundamental design choice: the network and the remote endpoint need source information so replies can find their way back.

Your IP address is an identifier. It connects to your approximate physical location and your internet service provider, which in turn connects to your legal identity through subscriber records. Every website you visit, every service you connect to, receives your IP address. They know where you are connecting from, and with additional records or account linkage they may infer who you are.

Even when content is encrypted, the IP header is not. HTTPS hides what you read on a website; it does not hide that you connected. Your ISP reliably sees that you are communicating and to which IP addresses or network providers; it often also sees the destination domain through DNS or TLS metadata, though encrypted DNS and ECH can reduce that visibility. Network observers along the path see source and destination IP addresses on every packet.

Metadata: The Full Picture

Metadata is data about data. For communications, metadata includes who (sender and recipient identifiers such as email addresses, phone numbers, and IP addresses), when (timestamps of communications), how long (duration of calls, size of messages), how often (frequency of communication between parties), and where (location data from devices). Content encryption hides what was said. Metadata reveals everything else.

Hayden's statement that "we kill people based on metadata" is not hyperbole. Communication patterns alone reveal social structure (who the leaders and intermediaries are), behavioral shifts (a sudden spike or silence signalling planning or operation), physical movement through device metadata, and association with known

targets. Intelligence agencies have acknowledged using this information for targeting decisions.

Why Content Encryption Is Insufficient

Consider encrypted messaging between two parties. An observer who cannot read the messages still sees that Alice and Bob communicate, how frequently they communicate, when they communicate (times of day, days of week), how their communication patterns change over time, and who else each party communicates with.

Such information suffices for surveillance purposes in many contexts. Knowing that a journalist communicates frequently with a particular government official is valuable intelligence regardless of message content.

Traffic Analysis

Traffic analysis is the systematic exploitation of metadata. Timing correlation observes that if Alice sends a message at 2:03:47 and Bob receives one at 2:03:48, they are probably communicating, even if the content is encrypted and the route is indirect. Volume correlation matches message sizes across network hops to link sender and receiver. Pattern analysis identifies regular communication patterns (every Tuesday at 3pm) that reveal relationships even without content. Network flow analysis follows traffic through network infrastructure to reveal endpoints even when individual hops are encrypted. Traffic analysis works because communication must traverse physical infrastructure that can be observed.

17.2 Requirements for Anonymous Communication

Formal Properties

Anonymous communication systems aim to provide several properties, each addressing different aspects of the metadata problem.

Sender anonymity means observers cannot determine who sent a message, so that even if the content is known, the originator remains hidden. This property protects whistleblowers and anyone whose speech might invite retaliation. Receiver anonymity means observers cannot determine who received a message. The intended audience is hidden, protecting recipients from association with senders who may be targeted.

Unlinkability means observers cannot determine that a particular sender and receiver are communicating with each other. Even if both parties are known to use the system, connecting their activity defeats surveillance that relies on mapping relationships. Unobservability means observers cannot determine that a communication is occurring at all. The communication is hidden among other traffic or cover activity. This is the strongest property because it hides the fact of communication itself.

These properties have different strengths depending on the adversary model. Against a local observer (seeing only part of the network), weaker protections may suffice. Against a global adversary (seeing all network traffic), stronger protections are required. The design choice reflects expected threats: journalists in authoritarian regimes face different adversaries than users seeking privacy from advertisers.

Adversary Models

The strength of an anonymity system depends on assumptions about the adversary it is designed against. A passive adversary observes traffic without modifying it, collecting data and attempting identification through pattern correlation. An active adversary can inject, delay, drop, or modify messages. They might run their own nodes in the network or perform timing attacks by introducing deliberate delays to force users into identifiable behavior.

The local adversary sees only a portion of the network: the user's connection to their ISP, or traffic through a single relay. The global adversary sees all network traffic simultaneously. This is the most powerful model, as it enables end-to-end timing correla-

tion that no amount of encryption can prevent without additional countermeasures.

Realistic threat modeling starts from who the likely adversary is. Most users do not face nation-state adversaries. But systems designed only for weak adversaries fail catastrophically when stronger ones appear. The cypherpunk approach builds systems that resist the strongest plausible adversaries, recognizing that capabilities expand over time.

The Anonymity Set

Anonymity is relative to an anonymity set: the group of possible senders or receivers among whom the actual party cannot be distinguished. This is the fundamental measure of anonymity strength.

If the anonymity set contains only three people, the adversary has a 1-in-3 chance of correct identification. If it contains millions, the odds improve correspondingly. The mathematics are simple: anonymity depends on who else is using the system.

Anonymity sets depend on who else is using the system at the same time. This creates a collective action dynamic: the more users, the stronger the anonymity for each user. A system with few users provides weak anonymity regardless of cryptographic strength. A system with many users can provide strong anonymity even with simpler cryptography.

This dynamic explains why adoption matters as much as technology. Privacy tools benefit from broad adoption. Early adopters sacrifice some anonymity to bootstrap the system; later adopters benefit from the anonymity set the pioneers created. The collective action problem also creates vulnerability: if adversaries can reduce adoption (through legal pressure or stigmatization), they weaken anonymity for remaining users.

Cover Traffic and Dummy Messages

Some systems introduce cover traffic, which is fake messages indistinguishable from real ones. Cover traffic serves several pur-

poses at once: it maintains consistent traffic volume regardless of actual usage, which defeats volume analysis, and it creates activity even when users are idle, making timing analysis harder. It expands the effective anonymity set by including dummy messages among possible "real" messages.

Cover traffic has costs. Bandwidth consumption increases, as dummy messages consume the same resources as real ones. Latency may increase if systems wait to batch real and dummy traffic. Complexity increases, since distinguishing cover from real traffic must be impossible for observers but possible for recipients.

The design choice depends on threat model. Against passive local adversaries, cover traffic may be unnecessary overhead. Against active global adversaries, it may be essential for effective protection.

17.3 VPNs: A Simple but Limited Solution

What VPNs Provide

VPNs (Virtual Private Networks) are the simplest approach to hiding an IP address. They encrypt traffic between the user and the VPN provider's server, which then forwards it to the destination. The local network segment is protected against observers such as coffee shop WiFi operators, the ISP cannot see final destinations even though it still sees the VPN connection itself, and destinations see the VPN server's IP instead of the user's. For many users, this is enough. If the concern is an ISP logging browsing history or a network operator sniffing traffic, a VPN addresses the immediate problem.

Trust Concentration and Its Consequences

A VPN does not eliminate surveillance; it relocates it. The VPN provider knows the user's real IP address and sees all traffic destinations. The surveillance has moved from the ISP to the VPN provider. If the provider logs traffic or cooperates with authorities, the user has no protection. Websites still see cookies and behavioral fingerprints that identify users regardless of the IP address

the packets carry. Unlike multi-hop systems, VPNs concentrate trust in a single commercial entity, and compromise of that entity compromises everything that flowed through it.

The trust model is also structurally opaque. VPN providers make claims about logging policies that users cannot verify. "No logs" claims have been contradicted when providers have turned over data to authorities, and users have no way to audit provider practices independently. Even well-intentioned providers can be compelled by legal process to begin logging, acquired by less privacy-respecting companies, or breached. The user must trust not only the provider's current stance but its future behavior under pressure it has not yet faced.

Some providers offer multi-hop configurations that route traffic through two or more servers, and users can chain VPNs manually by connecting to provider A and then through that connection to provider B. Provider A sees the real IP but not the destination; provider B sees the destination but only provider A's IP. Multi-hop VPNs remain weaker than Tor: the pool of providers is small compared to Tor's thousands of relays, the providers are commercial entities with known identities that are easier to pressure than pseudonymous relay operators, a single provider often controls multiple hops in its own "multi-hop" offering, and VPN configurations tend to be static where Tor circuits rotate. Single-hop VPN improves the path, but a purpose-built anonymity network distributes trust across a larger and less concentrated relay set.

Where the Trust Model Holds

The VPN trust model holds when the threat is local and the provider is outside the threat model. Protecting against ISP logging or public WiFi sniffing falls within that boundary, and for those threats a VPN does real work at low friction.

The model breaks when the adversary can reach the provider. Anonymity against a state that can subpoena the provider, protection from the provider itself, and any activity where concentrating trust in a single commercial entity is unacceptable all exceed what

VPNs can deliver. The architecture determines the boundary, and no choice of provider changes it.

17.4 Onion Routing: Tor and I2P

The Core Insight: Distribute Trust

The fundamental weakness of VPNs is that one entity sees everything. Onion routing solves this by distributing trust across multiple relays. No single relay knows both who you are and what you are accessing.

Tor: Architecture and Operation

Tor (The Onion Router) is the most widely used anonymous communication system.[2] It uses layered encryption where each layer is decrypted by successive relays.

The user's Tor client builds a circuit through three relays: a guard (entry) node, a middle node, and an exit node. The client knows all three; each relay knows only its neighbors. The message is encrypted three times, once for each relay; the guard decrypts the outer layer and forwards to middle, middle decrypts and forwards to exit, exit decrypts and forwards to destination. No individual relay knows both origin and destination: the guard knows the user but not the destination, the exit knows the destination but not the user, and the middle knows neither. This architecture means that one compromised relay does not automatically give an adversary a complete end-to-end view, though malicious relays still learn position-specific metadata. An adversary generally needs multiple vantage points or auxiliary attacks to link user to destination.

Tor Network Economics

Tor operates through volunteer relay operators who donate bandwidth and computing resources, and the incentive structure relies on self-interest combined with ideological commitment. Some operators need Tor themselves and contribute to the network they use; others operate relays to support freedom of communication as

a matter of principle. Some relays are operated by organizations such as universities and privacy advocacy groups as part of their institutional missions. This volunteer model creates sustainability challenges. Exit relays face particular burdens: abuse complaints and legal exposure. The network has consistently struggled with insufficient exit capacity.

Tor Limitations

Tor has well-documented limitations, and real-world attacks have showed these vulnerabilities.[3]

Timing attacks allow a global adversary observing both ends of a circuit to correlate timing and link sender to receiver; Tor does not protect against adversaries who can monitor traffic at both entry and exit points. Traffic analysis can reveal information from patterns in circuit usage even without breaking encryption. Website fingerprinting attacks analyze the size and timing patterns of encrypted traffic to identify which websites a user visits; research has achieved over 95% accuracy in controlled settings when monitoring a small set of popular websites, though real-world effectiveness remains debated.[4]

Exit node vulnerabilities exist because exit nodes see unencrypted traffic to destinations unless the destination uses HTTPS, allowing malicious exit operators to observe and modify unencrypted traffic. Guard node compromise is particularly serious because Tor users maintain the same guard nodes for extended periods; an adversary who controls a user's guard sees all their Tor traffic entering the network, and combined with exit observation or website fingerprinting, guard compromise can materially increase deanonymization risk.

Documented deanonymization attacks have succeeded against Tor users, though the Tor Project's ongoing maintenance has addressed many specific vulnerabilities. In Operation Torpedo (2012), the FBI deployed malware through compromised onion services to unmask users by exploiting browser vulnerabilities.[5] The 2013 Freedom Hosting attack used similar techniques.[6] Both attacks exploited browser plugins (particularly Flash) that Tor Browser now

disables by default; the specific vulnerabilities were patched. In 2014, researchers (allegedly from CMU/CERT) operated over 100 malicious relays comprising 6.4% of guard capacity, using a "relay early" traffic confirmation attack to deanonymize onion service users; The Tor Project discovered the attack and patched the "relay early" vulnerability in July 2014, ejecting the malicious relays.[7]

These historical attacks illustrate an important pattern: Tor undergoes continuous security review and improvement. Specific implementation vulnerabilities, when discovered, are typically patched promptly. The attacks that succeeded exploited bugs that no longer exist. What remains are structural limitations inherent to Tor's low-latency design: timing correlation by global adversaries, traffic analysis, and website fingerprinting cannot be fully eliminated without different architecture. Users should distinguish between historical exploits (largely fixed) and structural constraints (inherent to the design).

Nation-states have also showed sophisticated capabilities for detecting and blocking Tor bridges, the unlisted entry points designed to circumvent censorship. China, Iran, and Russia have implemented bridge-blocking with varying degrees of success. Sybil attacks, where an adversary creates many pseudonymous identities to gain disproportionate influence over a network, allow adversaries operating many relays to increase their chance of being selected for circuits, improving attack capabilities.

Performance suffers because multi-hop routing increases latency; Tor is slower than direct connections, sometimes substantially. Usability remains challenging; Tor is harder to use than ordinary browsing, and users make mistakes that compromise anonymity.

Tor's directory authority system is a point of centralization.[8] Ten directory authorities vote regularly to produce the network consensus document that lists all relays and their trustworthiness. Clients download this consensus to build circuits. The directory authorities themselves are known entities with stable identities, operated by trusted community members and organizations. While compromise of a single authority has limited impact due to the voting mechanism, the system as a whole depends on this small group remaining honest

and uncompromised. This is a pragmatic design choice: fully decentralized consensus is difficult for relay discovery, and the directory authority model has worked adequately. But it differs from the fully trustless models that some systems aspire to.

Tor provides strong protection against many adversaries, but well-resourced states can still disrupt access, block bridges, and improve their attack position by operating infrastructure or correlating traffic.

Onion Services: Censorship-Resistant Publishing

Tor's best-known use is anonymizing outbound connections: users access regular websites without revealing their identity. But Tor also enables the reverse: publishing services without revealing the server's location or requiring any registration with domain authorities.

An onion service generates a cryptographic key pair. For modern v3 onion services, the `.onion` address is derived from the service's ed25519 identity key material and encodes the public key, checksum, and version in base32. The service connects to the Tor network and establishes introduction points. Clients connect through the Tor network to these introduction points, then establish a rendezvous circuit. Neither client nor server reveals their IP address to the other or to any relay.

Onion services require no domain registration. Traditional websites require domain names purchased through registrars who enforce identity requirements and can revoke domains under legal pressure. Onion addresses derive from cryptographic keys. No registrar exists to pressure and no DNS seizure is possible. The address is self-authenticating: if you reach the service, you have reached the right service, verified by cryptography and not by certificate authorities.

The server location remains hidden: its IP address never appears in any connection, so adversaries cannot identify which server to raid or which jurisdiction's laws apply. A website can be published from anywhere and remain accessible as long as any path through the Tor network exists.

SecureDrop[9] lets sources submit documents without revealing their location to a news organization; censored publications maintain presence despite takedown orders; forums and markets operate without the jurisdictional exposure that destroyed centralized predecessors; and conventional services like Facebook and the BBC run .onion versions to reach users in censoring countries. Limitations remain: onion services are slower because of the multiple hops, the long random addresses are hard to communicate, and operational security failures can still unmask operators through other means.

I2P: A Different Architecture for Internal Services

I2P (Invisible Internet Project) uses similar principles to Tor but with different design goals and architectural choices.[10]

Garlic routing differs from onion routing in a significant way: garlic routing can bundle multiple message elements (called "cloves") into encrypted packets instead of sending every piece individually. In current I2P practice, that often includes one primary message plus control cloves. Even so, the format makes traffic analysis harder because an observer cannot easily distinguish which element corresponds to which communication stream.

I2P also uses unidirectional tunnels in place of Tor's bidirectional circuits. Each communication requires four tunnels: outbound and inbound for each party. Data sent through I2P takes one path to the destination and a different path for responses. This architectural choice makes observation more difficult because an adversary cannot assume the return path mirrors the outbound path.

I2P is a self-contained network that hosts hidden services (called "eepsites") accessible only within the network; users primarily communicate with other I2P users instead of anonymizing connections to external sites. Because traffic stays within I2P, there are no exit nodes with their associated vulnerabilities and abuse issues. The architecture is distributed: every I2P user also routes traffic for others, creating a more symmetric network than Tor's client-relay distinction.

I2P has its own security challenges. The network relies on a distributed database (the "netDB") maintained by floodfill

routers. Research in 2013 showed that Sybil attacks against floodfill routers could compromise the network; attackers who controlled sufficient floodfill peers could manipulate the database to enable deanonymization.[11] The I2P project responded by implementing mitigations including separating the netDB into multiple sub-databases and improving peer selection algorithms. Like Tor, I2P undergoes continuous development; discovered vulnerabilities are addressed through software updates. The smaller network size compared to Tor means fewer resources for security research, but the project maintains active development and responds to reported issues.

I2P's tradeoffs differ from Tor's. Fewer users means smaller anonymity sets. Tor has received much more academic scrutiny, leaving I2P's security properties less thoroughly analyzed. The focus on internal services makes accessing the regular internet less convenient than with Tor. I2P is appropriate for users whose primary need is communication with other I2P users and not anonymous access to the general internet.

17.5 Mixnets: The Strongest Protection

Why Onion Routing Is Not Enough

Tor and I2P protect against adversaries who cannot observe the entire network. But a global adversary, one who can monitor traffic entering and leaving the network simultaneously, can perform timing correlation. If a message enters Tor at 2:03:47.123 and exits at 2:03:47.456, the timing links them regardless of the encryption layers in between.

Mixnets address this limitation at the architectural level.

Chaum's Original Vision

David Chaum proposed mixnets in 1981, before Tor existed.[12] The concept: messages are collected, batched, reordered, and forwarded by mix nodes. Batching and reordering defeat timing

analysis by breaking the relationship between input and output timing.

How Mixing Defeats Traffic Analysis

A mix node accumulates messages over some time period. Once enough have arrived, the node collects them into a batch, strips the outer encryption layer from each, shuffles the batch into a new order, and forwards every message simultaneously.

An observer seeing messages enter and leave the mix cannot link inputs to outputs. Timing correlation fails because all outputs leave together. Order correlation fails because the order is shuffled. Even a global adversary who sees everything cannot determine which input corresponds to which output.

High Latency as Necessary Tradeoff

Strong mixing generally imposes latency and overhead. Messages often wait for batches or probabilistic delays to accumulate. Classic batch mixnets are poorly suited to interactive communication (instant messaging, web browsing) but suitable for asynchronous communication (email, file transfer, cryptocurrency transactions).

The tradeoff is fundamental: lower latency usually means weaker anonymity guarantees, while higher latency enables larger effective anonymity sets and stronger resistance to traffic analysis. Modern low-latency mixnets try to recover usability through cover traffic and different delay models, but they manage the tradeoff; they do not eliminate it.

Modern Implementations

Modern mixnet projects include Nym, which uses the Sphinx packet format[13] and economic incentives for mix operators.[14] Nym introduces cover traffic (fake messages) to further defeat traffic analysis and uses cryptocurrency-based incentives instead of volunteer

operation. Loopix is a mixnet design providing sender-receiver unlinkability with resistance to active attacks.

These projects attempt to make mixnets practical for modern use while preserving their strong anonymity properties. A critical limitation remains: deployed general-purpose mixnets still have far smaller user populations than Tor or mainstream VPN ecosystems. High-latency mixnets for email (like the historical Mixmaster and Mixminion systems[15]) are defunct. Modern mixnets like Nym remain in early deployment with limited adoption. The theoretical strength of mixing is real, but anonymity depends on who else is using the system; a mixnet with few users provides weaker practical anonymity regardless of cryptographic sophistication. For applications where latency is acceptable and adoption is sufficient, mixnets provide the strongest available protection against passive traffic analysis.

17.6 Why No Single Anonymity Architecture Suffices

The tradeoff between latency and anonymity strength is structural, not a design flaw. Low-latency systems preserve timing information that global adversaries can exploit; high-latency systems destroy timing information but cannot support interactive use. VPNs concentrate trust in a single provider for speed. Tor distributes trust across three relays, sacrificing latency so no single point of compromise reveals both endpoints. Mixnets destroy timing correlation through batching and reordering at the cost of interactivity. I2P optimizes for internal network services and accepts reduced external utility.

Each architecture answers a different adversary. A local observer is defeated by a VPN, a partial-path observer by Tor's multi-hop distribution, a global passive observer only by mixing. No universally optimal tool exists. Practical anonymity means matching a tool's known limitations to the actual threat. The application-layer protocols examined next ride on top of whichever transport the user chooses; payload encryption does not substitute for transport-layer

anonymity, and transport-layer anonymity does not substitute for end-to-end encryption of contents.

17.7 When the Internet Is Not the Network

Every transport the chapter has examined so far assumes a working internet. A privacy architecture that holds only under that assumption fails against the adversary who controls or removes the internet itself. The Ukraine, Gaza, Myanmar, and Iran shutdowns of the last decade are the ordinary operating conditions under which privacy most matters. Mesh and offline-first transports carry information when the internet is unavailable or is itself the adversary.

Meshtastic: LoRa at the Edge of the Spectrum

Meshtastic is an open-source mesh-networking protocol that runs on inexpensive LoRa radios. LoRa trades data rate for range, achieving hundreds of bits per second over several kilometers under favorable conditions, and handheld devices cost the price of a restaurant meal. Nodes relay for one another without central coordination; payloads are end-to-end encrypted; the network has no authoritative directory or administrative boundary. Bandwidth is tight enough that only short text messages are practical, and line-of-sight physics constrain range regardless of the protocol. An adversary who takes down the internet does not take down the spectrum, and the defender's fallback requires only that enough participants already hold the hardware.[16]

Reticulum and the Cryptographically Self-Routing Stack

Reticulum Network Stack (RNS) is a cryptographic networking stack that can run over any transport, including LoRa, packet radio, serial links, and ordinary TCP/IP, with every node's address its public key and no central routing authority. A privacy architecture that rides on the ordinary internet when that is available and on amateur radio when that is not, without changing its security model, is more resistant than an architecture tied to a specific transport. The

adversary who wants to deny the user connectivity must then deny every available transport simultaneously.[17]

Briar, BitChat, and Bluetooth-Scale Mesh

Briar synchronizes messages over Bluetooth, Wi-Fi, and Tor without a central server, targeting activists and journalists under compromised or absent networks. BitChat, released by Jack Dorsey in July 2025, runs over Bluetooth low energy mesh on iOS and Android with no phone number and no account, using store-and-forward relay so a message reaches its destination once any path of intermediate phones connects sender to receiver. Uptake on launch was rapid, and the category has expanded as users and journalists have carried these tools into the environments (protests, disasters, shutdowns) where the conventional network cannot be trusted. A Bluetooth mesh is the low-cost, low-range counterpart to a LoRa mesh; the defender's stack can include both at bounded cost.[18]

FIPS: A Self-Sovereign Network Layer

The Free Internetworking Peering System is a transport-agnostic overlay network in which every node generates its own cryptographic identity and routes without registration, DHCP, DNS, or any certificate authority.[19] A FIPS node's routing address is derived from a secp256k1 keypair the operator generates offline. Per-device network keys, rotatable independently of any other identity the operator holds, keep the peering graph from collapsing into the user's other trust domains.

What FIPS contributes to the transport stack is self-sovereign addressing paired with transport indifference. The same mesh can span UDP, TCP, Ethernet, Tor, and Bluetooth low energy simultaneously; a single packet can traverse any of them. Hop-by-hop links use Noise IK encryption so every peer-to-peer link is authenticated and confidential, and end-to-end sessions use Noise XK with periodic rekey so intermediate routers learn only 16-byte truncated node addresses, not payloads or endpoint identities. The spanning-tree construction is partition-tolerant: a local segment whose link to the

wider mesh is cut continues operating on its own and re-merges automatically on reconnect. An optional IPv6 adaptation layer maps node addresses onto a `fd00::/8` range so unmodified applications can address any FIPS endpoint by hostname. Where Tor provides low-latency anonymity over the existing internet, FIPS provides addressing and routing that do not require the existing internet at all.

Amateur Radio and Licensed Spectrum

Licensed amateur radio is the oldest mesh. Operators relay messages using protocols including Winlink, FT8, JS8Call, and APRS, reaching everywhere an antenna can be pointed and a license obtained. FCC Part 97 in the U.S. (with analogues elsewhere) prohibits obscuring the meaning of messages, which is interpreted as a prohibition on encryption. Amateur radio can therefore carry authenticated but not confidential traffic in most jurisdictions. A signed Nostr event transmitted over an amateur band remains cryptographically authentic even though its contents are readable, and under state-shutdown conditions the defender takes what the regulatory regime allows.[20]

Satellite and the Question of Sovereignty

Low-earth-orbit constellations (Starlink, Iridium, and competitors) provide connectivity that does not depend on the terrestrial network of any specific state. Starlink has been subject to documented geofencing under political pressure; older networks like Iridium have had less bandwidth and less political attention and have proven more stable as a consequence. Satellite links belong in a privacy stack because they are orthogonal to terrestrial internet; which operator the defender chooses depends on specific political exposures.[21]

The Common Thread

Every transport in this section answers the same question: when the conventional network is unavailable or hostile, what carries the message? The answers differ by range, bandwidth, regulatory status, and political exposure. What they share is that each is cheap to deploy at the node, expensive to suppress at the network, and scalable in density as threat levels rise. Encryption protects what is said; mesh and offline-first transports protect whether it can be said at all.

17.8 Encrypted Messaging: The Application Layer

The transport-layer tools in the preceding sections hide the fact of communication. Encrypted-messaging protocols hide its contents. The two sit at different layers of the stack and solve different problems, but their histories are entangled: each generation of messaging protocol has tried to protect more of what the last one left exposed, and the residual leaks motivate exactly the transport-layer tools this chapter has already examined.

PGP: Content Encryption and Its Limits

Pretty Good Privacy, released by Phil Zimmermann in 1991 and standardized as OpenPGP in RFC 4880, was the first widely available tool that let ordinary people encrypt and sign email.[22] Its arrival was politically significant: Zimmermann was investigated for three years under U.S. arms-export regulations on the theory that his software constituted a munition.[23] The investigation was dropped, but it established that civilian access to strong cryptography would be contested. The chapter on the cypherpunk movement returns to this episode; the protocol that came out of it shaped the next generation of message encryption.

Cryptographically, PGP is a hybrid scheme. A long-lived primary key, conventionally split into a signing primary and one or more encryption and authentication subkeys, identifies the user. To send a message, the sender generates a fresh symmetric session

key, encrypts the payload with that session key under a symmetric cipher, and then encrypts the session key separately to each recipient's public encryption subkey. Signatures over the plaintext or ciphertext are produced with the sender's signing primary or a signing subkey. The use of subkeys lets users rotate operational keys without abandoning the long-term identity that already accumulated trust signatures, but in practice most users never rotate at all.

The trust model is the web of trust. There is no certificate authority. Users sign each other's public keys after verifying them in person, and trust propagates transitively along signed chains. Public keys are published to keyservers, and verification is supposed to happen out of band, often through key signing parties at conferences where participants check government identification and exchange key fingerprints. The model is intellectually elegant and operationally hostile to anyone outside a small technical community.

What PGP does well is also what limits it. Long-lived keys are useful for signing software releases, Git commits, package repositories, and other artifacts whose authorship should remain provable for years; this is why PGP survives in the supply-chain integrity role even where it has been abandoned for messaging. The same long life is a liability for live correspondence. PGP provides no forward secrecy: anyone who compromises a long-term private key, whether through device seizure, malware, a cold subpoena, or a quiet decryption advance, can decrypt every past message ever encrypted to that key. Signatures are non-repudiable by design, which is the desired property for software releases and the wrong property for personal conversation; an offhand remark sent to one person becomes durable cryptographic evidence of authorship that the sender cannot later deny.

The trust infrastructure also failed in practice. Most users never participated in key signing and could not meaningfully verify the keys they downloaded. The SKS keyserver pool, which served as the de facto key distribution system for years, had no abuse controls; an attacker discovered in 2019 that flooding a key with hundreds of thousands of meaningless signatures produced a key file so large

that GnuPG would hang or crash when it tried to import it, and the keys of high-profile project maintainers were poisoned faster than they could be rotated. The keyserver network never recovered. Newer distribution mechanisms exist, including the Web Key Directory standard that lets a domain publish keys at a well-known URL under its own authority, the Hagrid keyserver run by the Sequoia project under stricter validation rules, and the Autocrypt opportunistic key exchange built into clients like Thunderbird and K-9 Mail. None has restored a single working substitute for what SKS was supposed to do, and the modern OpenPGP world is fragmented across competing implementations such as GnuPG, Sequoia, RNP, and OpenPGP.js, with the LibrePGP fork having split from RFC 9580 in 2024.

Usability has been a known problem for as long as PGP has existed. The 1999 paper "Why Johnny Can't Encrypt" tested PGP 5.0 with educated users and found that most could not reliably encrypt or decrypt mail without making mistakes that defeated the protection. Later work in 2015 found the same problems unfixed in modern Mailvelope and similar tools. The interface model never matched the conceptual model, and the conceptual model never matched what users needed.

PGP also encrypts content only. Email headers, subject lines, recipient addresses, timestamps, and the key identifiers in encrypted message packets all travel in the clear, which lets a passive observer reconstruct the entire social graph and timing pattern of supposedly private correspondence. For secure messaging, PGP solves one problem and leaves many others visible. The protocols that followed treated each of those exposed surfaces as a problem to address explicitly.

OTR: Forward Secrecy and Deniability for Live Chat

Off-the-Record Messaging, introduced by Nikita Borisov, Ian Goldberg, and Eric Brewer in 2004, was a deliberate answer to PGP's model.[24] Their paper is direct about the target: PGP-style cryptography makes live conversation look nothing like live conversation. Ordinary speech is authenticated in the moment but leaves

no durable proof afterward, and the speakers can later deny what they said with no cryptographic instrument standing against them. OTR asked what cryptography would look like if it behaved the same way, and built the answer on top of existing instant messaging carriers, principally XMPP, AIM, ICQ, and IRC.

The protocol provides four properties: encryption, authentication, perfect forward secrecy, and deniability. Authentication runs on a per-session Diffie-Hellman exchange signed once at session setup with long-term identity keys, so parties prove who they are at the start of the conversation without re-signing every message. Message integrity runs on message authentication codes instead of digital signatures, because MAC keys are symmetric and shared between the two participants; either side could have generated any given MAC, so neither MAC alone is evidence against the other. After a MAC key is no longer needed for verification, OTR publishes it on the wire, making transcripts retroactively forgeable by anyone who later obtains them. Diffie-Hellman keypairs ratchet forward through the session as new key material is exchanged, so compromise of any single ephemeral key compromises only a small window of messages.

A second deniability problem OTR addressed is identity authentication itself. Asking "is this Alice?" over an untrusted channel is hard if the answer must be unforgeable to a network attacker but the channel is the only one available. OTR solved this with the Socialist Millionaire Protocol,[25] an interactive zero-knowledge construction that lets two parties confirm they share a secret answer to an agreed question without revealing the answer to anyone, including each other if they got it wrong. In practice this allows two people who already know something private together (a shared past, a phrase agreed in person) to verify each other's identity without exchanging long key fingerprints and without producing a signature anyone could replay.

OTR's contribution to the line was conceptual: it showed that forward secrecy and deniable authentication were simultaneously achievable for interactive use, and that the property set ordinary conversation already has could be reproduced cryptographically.

The protocol shipped initially as a libpurple plugin for Pidgin and Adium[26] and was later integrated into clients including Jitsi, ChatSecure, CryptoCat, and Tor Messenger. OTR version 3, published in 2012, added multi-instance support so a user could run OTR from more than one client at once, and OTRv4, drafted from 2017 onward by the OTR Development Team, replaced the original key exchange with a deniable authenticated key exchange and added a double-ratchet construction inspired by the work that became Signal.[27]

The decline was operational. OTR assumed both parties were online at the same time to run the interactive key exchange, and it assumed a single long-running session per pair of devices. The smartphone era broke both assumptions. Mobile clients went offline whenever a screen locked or network connectivity hiccuped, and users acquired multiple devices that they expected to share a single conversation. OTR never acquired native asynchronous or multi-device support that worked at the level mobile users had come to expect from iMessage and WhatsApp, and the underlying XMPP networks that hosted most OTR conversations themselves shrank as users moved to integrated platforms. The successor protocol absorbed OTR's ratcheting idea into an architecture designed for the new constraints.

Signal: Asynchronous Forward Secrecy at Smartphone Scale

The Signal Protocol, developed by Trevor Perrin and Moxie Marlinspike at Open Whisper Systems, is the engine behind Signal itself and behind the end-to-end encryption shipped by WhatsApp, Google Messages, Meta Messenger, and most of the other mainstream messengers that claim E2EE.[28][29] It is two distinct protocols working in sequence.

X3DH completes a key agreement when the other party is offline by combining pre-published keys, and the Double Ratchet then rotates keys per message to provide both forward secrecy (past messages remain protected if the current state is compromised) and post-compromise security (future messages become protected again after the attacker is locked out for one round-trip). Out-of-order

delivery is handled by caching skipped message keys in bounded windows.

Signal has been deployed at extraordinary scale and shipped post-quantum key agreement (PQXDH) in 2023. What it does not solve natively is group messaging. Signal groups are implemented with "sender keys": each participant broadcasts a symmetric ratchet to the others over their pairwise Double Ratchet channels. Adding or removing a member, or rotating keys after compromise, costs O(N) pairwise operations, which limits practical group sizes and makes post-compromise security expensive at scale. Signal also does not hide who is talking to whom; sealed-sender and private contact discovery are application-layer additions on top of a protocol that still requires a central server to deliver messages. The Signal server is a single shutdown target.

Noise: A Framework for Secure Channels

Signal and MLS solve the messaging-application problem. The Noise Protocol Framework, published by Trevor Perrin in 2016, solves the layer below: how to build a two-party authenticated and confidential channel in the first place, without relying on the certificate hierarchy that TLS assumes. Noise is not a protocol but a construction kit. A designer picks a handshake pattern (IK, XK, XX, and others are named by their authentication shapes), a Diffie-Hellman curve (25519 or 448), a cipher (ChaCha20-Poly1305 or AES-GCM), and a hash (SHA-256, SHA-512, or BLAKE2), and the framework derives a concrete protocol whose properties are specified by the chosen pattern. Forward secrecy, mutual authentication, and optional identity hiding are pattern-level guarantees, not implementation details.

The patterns carry their properties in the name. IK means the initiator knows the responder's static key in advance, so the initiator can authenticate the responder on the first message; XK is the same pattern with initiator identity hidden until the responder proves itself, which is useful when the initiator wants to stay anonymous to passive observers. XX is mutual authentication without either party knowing the other's static key in advance, which is the pattern TLS

1.3 approximates. Each pattern's security properties are formally analyzed, and a handshake built from a named pattern inherits those properties by construction.

WireGuard uses Noise_IK_25519_ChaChaPoly_BLAKE2s for its VPN handshake, which is why a WireGuard peer establishes an encrypted tunnel in one round-trip and the handshake format is small enough to fit in a UDP packet. Signal's X3DH and the I2P NTCP2 transport are Noise-derived. Lightning's BOLT #8 transport encryption uses Noise_XK for node-to-node communication, so every Lightning channel establishment includes a mutually authenticated handshake with responder-identity privacy. FIPS uses Noise_IK for hop-to-hop links and Noise_XK for session-layer encryption, inheriting both the handshake brevity and the formal analysis that comes with the named patterns. The framework has become the default construction for new secure-channel designs because the design work has already been done; a protocol author picks the pattern whose properties match the adversary model and gets a proved construction without writing a new one.

MLS: Group Messaging as a Standard

Messaging Layer Security, standardized as RFC 9420 by the IETF in 2023, is the first standards-track answer to the group-scaling problem Signal's design could not efficiently solve.[30] MLS is a continuous group key agreement protocol: an authenticated key exchange designed from the start for groups instead of pairs, scaling from two members up to the tens of thousands the RFC targets as its working range.

MLS uses a ratchet tree, reducing the cost of adding, removing, or updating a member to $O(\log N)$ operations from Signal's $O(N)$ pairwise sender-keys. Group lifetime advances through explicit epoch transitions that inject fresh entropy, and new members receive the group state asynchronously via prekey-style KeyPackages.

The security properties MLS inherits from this structure are exactly what Signal provides pairwise, lifted to groups: forward secrecy through epoch rotation, and post-compromise security because any commit by or about a compromised member introduces

entropy the attacker does not control. What MLS leaves to the surrounding system is metadata protection. The RFC explicitly notes that the delivery service sees group identifiers, message frequency, and member counts. Hiding the social graph around the messages is not MLS's job. Cisco Webex, Wire, RingCentral, and Matrix have deployed MLS in production, and the GSMA's Universal Profile 3.0 specifies MLS for cross-carrier interoperable end-to-end encryption in RCS,[31] with corresponding plans announced for iMessage. The IETF MIMI working group is building on MLS to satisfy European interoperability requirements.[32]

A separate line of deployment composes MLS with the decentralized relay layer examined in Chapter 21. The Marmot protocol uses Nostr keys as the MLS Authentication Service and Nostr relays as the MLS Delivery Service, so the identity and transport pieces MLS deliberately leaves abstract are filled in by infrastructure that no single entity controls. Marmot also addresses the metadata gap the RFC concedes. Each group message is published under a fresh ephemeral relay-layer keypair instead of the user's identity key, and the public group identifier is a rotatable pseudo-ID instead of the real MLS group ID. Welcome messages are wrapped in onion-style gift-wrap envelopes that hide sender and recipient from the relays carrying them. White Noise, the flagship Marmot client, ships on Android and iOS devices, and a specification audit by Least Authority completed in early 2026.[33] Chapter 21 treats this composition in detail; the point here is that MLS's deliberate separation of cryptographic core from authentication and delivery is what makes such a composition possible at all.

What the Progression Shows

Each generation solved the core problem the previous one left open and revealed the next one. PGP protected content and exposed everything else. OTR protected live conversation but assumed synchronous single-device use. Signal made strong E2EE asynchronous and ubiquitous for pairs but scaled poorly to groups and depended on a central server. MLS standardized efficient group encryption but left the metadata problem at the delivery service to the surrounding

system. That surrounding system is exactly what the transport-layer tools in sections 17.3 through 17.5 try to address: even the strongest application-layer encryption leaves the delivery infrastructure with the power to observe patterns of communication, so the two layers compose only when both are present. Chapter 21 examines how a decentralized relay layer can be combined with MLS to close more of that gap.[34]

Chapter Summary

Metadata exposure is a structural problem that content encryption cannot solve. The internet's design embeds sender and receiver identifiers in every packet, and encrypting the payload leaves communication patterns, timing, frequency, and volume visible to any observer along the path. Anonymous communication architectures form a spectrum defined by the latency-anonymity tradeoff. VPNs relocate trust to a single provider in exchange for speed; onion routing distributes trust across multiple relays so no single node learns both endpoints; mixnets destroy timing correlation through batching and reordering at the cost of latency that rules out interactive use. The adversary model determines the appropriate tool. A local observer is defeated by a VPN, a partial-path observer by Tor's multi-hop distribution, a global passive observer only by mixing. When the conventional internet is itself the adversary, mesh and offline-first transports (Meshtastic, Reticulum, Briar, BitChat, FIPS, amateur radio, satellite) carry messages on infrastructure no single state controls. A privacy architecture benefits from composing several of these transports so that an adversary who wants to deny connectivity must deny every available channel at once.

Application-layer protocols have progressively strengthened what encryption protects. PGP encrypted content and left everything else exposed. OTR added forward secrecy and deniability for live chat. Signal made asynchronous forward-secret messaging practical at smartphone scale. Noise became the default construction for new secure channels. MLS provides the same Signal-style security guarantees for groups in O(log N) operations and shipped as an

IETF standard in 2023. Each generation solved what the previous left open and left the metadata problem at the delivery service to the layers below. No single tool provides complete anonymity, and every architecture has structural limitations determined by its design tradeoffs. Chapter 22 addresses the operational security practices that complement architectural protection, and Chapter 20 examines the distinct metadata problems Bitcoin's on-chain transparency and Lightning's channel graph create for monetary privacy.

Notes

[1]The quote is widely attributed to Michael Hayden from a 2014 debate at Johns Hopkins University. See David Cole, "We Kill People Based on Metadata," *New York Review of Books*, May 10, 2014. For the earlier structural statement of the same dossier problem in transaction systems, see David Chaum, "Security Without Identification: Card Computers to Make Big Brother Obsolete," *Communications of the ACM* 28, no. 10 (1985): 1030-1044, archived at `https://nakamotoinstitute.org/library/security-without-identification/`.

[2]Roger Dingledine, Nick Mathewson, and Paul Syverson, "Tor: The Second-Generation Onion Router," *Proceedings of the 13th USENIX Security Symposium* (2004): 303-320. The Tor Project, `https://www.torproject.org/`.

[3]For a full survey of attacks on Tor, see the "Attacks on Tor" repository maintained by security researchers, available at `https://github.com/Attacks-on-Tor/Attacks-on-Tor`.

[4]On website fingerprinting, see Giovanni Cherubin et al., "Online Website Fingerprinting: Evaluating Website Fingerprinting Attacks on Tor in the Real World," *USENIX Security Symposium* (2022). The study found that accuracy degrades rapidly when monitoring larger sets of websites in realistic conditions.

[5]On Operation Torpedo, see Kevin Poulsen, "FBI Used Drive-By-Downloads to Expose Tor Pedophiles," *Infosecurity Magazine*, August 2014. The NIT (Network Investigative Technique) exploited a Flash vulnerability to reveal real IP addresses.

[6]The 2013 Freedom Hosting operation was an FBI action that deployed malware (a Network Investigative Technique) through a JavaScript exploit against Firefox 17 ESR to unmask users of Tor hidden services hosted on Freedom Hosting's infrastructure. Kevin Poulsen, "FBI Allegedly Ran Tor Hidden Services, Seized Them to Unmask Child Porn Viewers," *Wired*, August 5, 2013. Subsequent analysis by Vlad Tsyrklevich, "Tor Browser Bundle Exploit Analysis" (August 2013), documented the payload's mechanism. The exploit targeted a Firefox use-after-free vulnerability (CVE-2013-1690) that had been patched in newer

Firefox versions but remained present in the then-current Tor Browser Bundle; the attack was mitigated by upgrading Tor Browser and disabling JavaScript.

[7]The Tor Project, "Tor Security Advisory: 'Relay Early' Traffic Confirmation Attack," July 30, 2014, available at `https://blog.torproject.org/tor-security-advisory-relay-early-traffic-confirmation-attack/`. See also Philipp Winter et al., "Identifying and Characterizing Sybils in the Tor Network," *USENIX Security Symposium* (2016).

[8]Tor's directory authority system and its consensus mechanism are documented in the Tor design paper and the dir-spec. Roger Dingledine and Nick Mathewson, "Tor Protocol Specification," at `https://spec.torproject.org/`; the directory authority protocol specifically in "Tor directory protocol, version 3," `https://spec.torproject.org/dir-spec`. The nine currently recognized directory authorities and their operators are listed at `https://metrics.torproject.org/rs.html#search/flag:Authority`. On the security properties and known weaknesses of the directory authority system, see Aaron Johnson et al., "Users Get Routed: Traffic Correlation on Tor by Realistic Adversaries," *ACM CCS* (2013), `https://www.ohmygodel.com/publications/usersrouted-ccs13.pdf`.

[9]SecureDrop is the open-source whistleblower submission system developed by the Freedom of the Press Foundation, built on Tor onion services; it is deployed by more than 75 news organizations. Freedom of the Press Foundation, `https://freedom.press/`. SecureDrop project at `https://securedrop.org/`, with source at `https://github.com/freedomofpress/securedrop`. The original design was created by Aaron Swartz and Kevin Poulsen as DeadDrop; the Freedom of the Press Foundation assumed stewardship in 2013. On SecureDrop's threat model and architecture, see the official documentation at `https://docs.securedrop.org/en/stable/threat_model/threat_model.html`.

[10]For I2P technical documentation, see `https://geti2p.net/en/docs`. On garlic routing specifically, see `https://geti2p.net/en/docs/how/garlic-routing`.

[11]Christoph Egger et al., "Practical Attacks Against the I2P Network," *RAID Symposium* (2013). The researchers showed successful attacks using only 20 malicious floodfill peers. For I2P's threat model and mitigations, see `https://geti2p.net/en/docs/how/threat-model`.

[12]David Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms," *Communications of the ACM* 24, no. 2 (1981): 84-90.

[13]On the Sphinx packet format used by modern mixnets, see George Danezis and Ian Goldberg, "Sphinx: A Compact and Provably Secure Mix Format," *IEEE Symposium on Security and Privacy* (2009), `https://cypherpunks.ca/~iang/pubs/Sphinx_Oakland09.pdf`. For the Loopix design that Nym partially descends from, see Ania M. Piotrowska et al., "The Loopix Anonymity System," *USENIX Security Symposium* (2017).

[14]Claudia Diaz, Harry Halpin, and Aggelos Kiayias, "The Nym Network," available at https://nym.com/.

[15]Mixmaster and Mixminion were the two deployed high-latency anonymous remailer networks for email; both are now defunct. Mixmaster was the type-II remailer network; Lance Cottrell, "Mixmaster and Remailer Attacks," in *Proceedings of the Workshop on Privacy Enhancing Technologies* (1994). Mixminion was the type-III successor: George Danezis, Roger Dingledine, and Nick Mathewson, "Mixminion: Design of a Type III Anonymous Remailer Protocol," *IEEE Symposium on Security and Privacy* (2003), https://mixminion.net/minion-design.pdf. Both networks closed due to insufficient user adoption rather than cryptographic failure, illustrating the collective-action dynamics the chapter describes. The Mixminion source archive remains at https://mixminion.net/.

[16]Meshtastic is the dominant open-source LoRa mesh protocol; Heltec, LILYGO, and RAK Wireless are the three main hardware vendors whose devices run it. Meshtastic project, https://meshtastic.org/, with source at https://github.com/meshtastic. LoRa modulation specification maintained by the LoRa Alliance at https://lora-alliance.org/. Hardware: Heltec Automation, https://heltec.org/; LILYGO, https://www.lilygo.cc/; RAK Wireless, https://www.rakwireless.com/. Event-scale deployment reporting in security-conference community coverage at DEF CON and similar gatherings since 2022.

[17]Reticulum is a transport-agnostic cryptographic networking stack where every node's address is its public key; NomadNet and LXMF are the messaging applications built on top of it. Reticulum Network Stack, https://reticulum.network/, documentation at https://markqvist.github.io/Reticulum/manual/. NomadNet, https://github.com/markqvist/NomadNet. LXMF (Lightweight Extensible Message Format), https://github.com/markqvist/LXMF.

[18]Briar and BitChat are the two reference Bluetooth-mesh messengers; Briar is the activist/journalist tool with Tor synchronization and a decade of peer-review history, BitChat (launched July 2025 by a project associated with Jack Dorsey) is the mass-market entry with no-account onboarding and iOS/Android distribution. Briar, https://briarproject.org/, with source at https://code.briarproject.org/briar/briar. On Briar's design and threat model, see Michael Rogers and Eleanor Saitta, "Briar: Private Messaging, Local Networks, and Censorship Resistance" (2018), https://briarproject.org/publications/. BitChat, https://bitchat.free/, with source at https://github.com/permissionlesstech/bitchat. On the security-research pushback against unaudited Bluetooth-mesh designs, see Matthew Green and Trail of Bits audit reporting at https://blog.cryptographyengineering.com/.

[19]FIPS (Free Internetworking Peering System) is a self-organizing transport-agnostic overlay network in which each node generates its own secp256k1 keypair as its routing identity, independent of any ISP, regional internet registry, or certificate authority. FIPS, `https://github.com/jmcorgan/fips`, with protocol design documentation at `https://github.com/jmcorgan/fips/tree/master/docs/design`. Released under MIT license; at v0.2.0 at time of writing. The Noise protocol framework, `https://noiseprotocol.org/`, underlies both the hop-to-hop (Noise IK) and session-layer (Noise XK) encryption. An optional IPv6 adaptation layer maps node addresses onto an `fd00::/8` range so unmodified applications can address any FIPS endpoint.

[20]U.S. FCC Part 97 and analogous regulations in most jurisdictions prohibit encryption on amateur bands; Winlink, FT8, JS8Call, and APRS are the deployed data modes; amateur radio carries authenticated but not confidential traffic under these regimes. FCC Part 97 rules at `https://www.ecfr.gov/current/title-47/chapter-I/subchapter-D/part-97`. Winlink, `https://www.winlink.org/`. FT8 (part of the WSJT-X software suite), `https://wsjt.sourceforge.io/`. JS8Call, `http://js8call.com/`. APRS (Automatic Packet Reporting System), `http://www.aprs.org/`. On signed Nostr events carried over authenticated but unencrypted amateur bands, see community reporting from the HamRadio + Nostr intersection at `https://github.com/search?q=nostr+ham+radio`.

[21]Starlink, Iridium, OneWeb, and Globalstar operate low-earth-orbit constellations that carry traffic independent of any state's terrestrial infrastructure; Starlink's geofencing and combat-zone disruptions make the case that no single operator is politically trustworthy and that orthogonal transport in the defensive stack beats reliance on any one satellite link. Starlink, `https://www.starlink.com/`. Iridium, `https://www.iridium.com/`. OneWeb (operated by Eutelsat Group), `https://oneweb.net/`. Globalstar, `https://www.globalstar.com/`. On Starlink's documented disruptions in Ukraine, see reporting in *Financial Times*, *Reuters*, and Elon Musk's own public statements on X/Twitter archived at `https://web.archive.org/`. On the geofencing incidents, see Ukrainian Ministry of Defense statements through 2023–2024.

[22]J. Callas et al., "OpenPGP Message Format," RFC 4880 (November 2007); updated by RFC 9580 (July 2024) and forked by the LibrePGP specification. On PGP's usability and architectural limits, see Alma Whitten and J. D. Tygar, "Why Johnny Can't Encrypt," *USENIX Security Symposium* (1999); Scott Ruoti et al., "Why Johnny Still, Still Can't Encrypt: Evaluating the Usability of a Modern PGP Client," *arXiv:1510.08555* (2015); and Matthew Green, "What's the Matter with PGP?" (2014), available at `https://blog.cryptographyengineering.com`. On the SKS keyserver poisoning attack of 2019, see Daniel Kahn Gillmor, "OpenPGP Certificate Flooding," and Robert J. Hansen, "SKS Keyserver Network Under Attack" (June 2019), the latter at `https://gist.github.com/rjhansen/67ab921ffb4084c865b3618d6955275f`.

On the Web Key Directory standard, see the IETF draft at `https://datatracker.ietf.org/doc/draft-koch-openpgp-webkey-service/`. On Autocrypt, see `https://autocrypt.org`.

[23]The U.S. government's arms-export investigation of Phil Zimmermann under the International Traffic in Arms Regulations (ITAR) ran from 1993 to 1996 and established that civilian access to strong cryptography would be legally contested. The case was dropped without charges, but its chilling effect on cryptography export shaped the cypherpunk movement's urgency. Philip R. Zimmermann, *The Official PGP User's Guide* (MIT Press, 1995), documents PGP's design and the political context. The ITAR case is documented in Simson Garfinkel, *PGP: Pretty Good Privacy* (O'Reilly, 1994), pp. 81–102, and in Steven Levy, *Crypto: How the Code Rebels Beat the Government Saving Privacy in the Digital Age* (Viking, 2001), pp. 199–230. Phil Zimmermann's own account appears at `https://web.archive.org/web/20240101000000/https://philzimmermann.com/EN/background/index.html`.

[24]Nikita Borisov, Ian Goldberg, and Eric Brewer, "Off-the-Record Communication, or, Why Not To Use PGP," *Workshop on Privacy in the Electronic Society* (2004). Specification at `https://otr.cypherpunks.ca/`.

[25]The Socialist Millionaire Protocol (SMP) is the zero-knowledge equality-test used in OTR for mutual identity authentication without exchanging long key fingerprints. It derives from Fabrice Boudot, Berry Schoenmakers, and Jacques Traoré, "A Fair and Efficient Solution to the Socialist Millionaires' Problem," *Discrete Applied Mathematics* 111, no. 1–2 (2001): 23–36. OTR's adaptation of SMP is specified in the OTR protocol version 2 and later specifications at `https://otr.cypherpunks.ca/Protocol-v3-4.1.1.html`; the specific implementation as an interactive zero-knowledge proof that two parties share a common secret without revealing it appears in section 2.6 of that document.

[26]Pidgin is the multiprotocol instant messaging client whose libpurple plugin architecture enabled OTR's initial distribution; Adium is the macOS equivalent. The original OTR libpurple plugin at `https://otr.cypherpunks.ca/`. Pidgin project at `https://www.pidgin.im/`. Adium project at `https://adium.im/`. The libpurple library, shared between Pidgin and Finch, provided the plugin API that allowed OTR to be retrofitted onto AIM, ICQ, XMPP, and other protocols without modifying the underlying clients.

[27]OTRv4 replaces the original Diffie-Hellman key exchange with a deniable authenticated key exchange (DAKE) based on the work of Yehuda Lindell and adds a double-ratchet mechanism that provides forward secrecy at the message level within a session. Specification: Nik Unger, Sofía Celi, Jurre van Bergen, and the OTR Development Team, "OTRv4," draft specification (2019–present), `https://github.com/otrv4/otrv4/blob/master/otrv4.md`. On the deniable authenticated key exchange construction, see Mario Di Raimondo, Rosario Gennaro, and Hugo Krawczyk, "Deniable Authentication and Key Exchange," *ACM CCS* (2006), `https://eprint.iacr.org/2006/280`.

[28]For the Signal Protocol specifications, see Moxie Marlinspike and Trevor Perrin, "The X3DH Key Agreement Protocol" (2016) and "The Double Ratchet

Algorithm" (2016), both at https://signal.org/docs/. For formal analysis, see Katriel Cohn-Gordon et al., "A Formal Security Analysis of the Signal Messaging Protocol," *IEEE European Symposium on Security and Privacy* (2017). Post-quantum extension: "The PQXDH Key Agreement Protocol," 2023. The Noise Protocol Framework specification, Trevor Perrin, "The Noise Protocol Framework," revision 34 (2018), https://noiseprotocol.org/noise.html, is the reference for the handshake-pattern taxonomy. Noise-derived deployments include WireGuard (Jason Donenfeld, "WireGuard: Next Generation Kernel Network Tunnel," *NDSS* 2017), Lightning's BOLT #8 transport encryption, I2P's NTCP2 transport, and FIPS's hop and session encryption.

[29]WhatsApp's 2016 deployment of the Signal Protocol to its then 1 billion users was the largest single rollout of end-to-end encryption in history. Moxie Marlinspike and Trevor Perrin, "WhatsApp's Signal Protocol Integration Is Now Complete," Open Whisper Systems Blog (April 5, 2016), https://signal.org/blog/whatsapp-complete/. The technical white paper is WhatsApp Inc., "WhatsApp Encryption Overview: Technical White Paper" (2016, updated 2017), https://www.whatsapp.com/security/WhatsApp-Security-Whitepaper.pdf. Google Messages' RCS end-to-end encryption using the Signal Protocol is documented at https://support.google.com/messages/answer/10262381. Meta Messenger's adoption is described in Jon Millican, "Messenger End-to-End Encryption Overview," Meta Engineering Blog (2023), https://engineering.fb.com/2023/12/06/security/messenger-end-to-end-encryption-default/.

[30]R. Barnes et al., "The Messaging Layer Security (MLS) Protocol," RFC 9420 (July 2023); R. Barnes et al., "The Messaging Layer Security (MLS) Architecture," RFC 9750 (2024). The underlying ratchet-tree design originates in Katriel Cohn-Gordon et al., "On Ends-to-Ends Encryption: Asynchronous Group Messaging with Strong Security Guarantees," *ACM CCS* (2018).

[31]The GSMA's Rich Communication Services (RCS) Universal Profile 2.4 introduced end-to-end encryption, and Universal Profile 3.0 specifies MLS as the group key agreement protocol, enabling cross-carrier interoperable E2EE in carrier-delivered messaging. GSMA, "RCS Universal Profile Service Definition Document," version 3.0 (2024), available to GSMA members at https://www.gsma.com/solutions-and-impact/technologies/networks/rcs/. Apple's announcement of RCS support with end-to-end encryption for iMessage interoperability: Apple Inc., "Apple Announces RCS Support" (2024), with technical details in the WWDC 2024 session "What's new in Messages for Business." Google's Messages implementation of RCS E2EE is documented at https://support.google.com/messages/answer/10262381.

[32]The IETF More Instant Messaging Interoperability (MIMI) working group is developing standards to allow end-to-end encrypted messaging across different providers, primarily to satisfy the European Union's Digital Markets Act interoperability requirements. IETF MIMI working group charter and documents at https://datatracker.ietf.org/wg/mimi/about/. The key documents

are the MIMI content format (draft-ietf-mimi-content) and the MIMI protocol (draft-ietf-mimi-protocol), both building on MLS (RFC 9420) as the group key agreement layer. On the Digital Markets Act interoperability obligation that motivates this work, see Regulation (EU) 2022/1925 of the European Parliament and of the Council, Article 7, `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R1925`.

[33]The Marmot protocol specification and White Noise client are the primary open-source implementation of MLS over Nostr infrastructure. Marmot protocol specification at `https://github.com/marmot-protocol/marmot`. White Noise client (Android and iOS) at `https://whitenoise.chat/`. The Least Authority security audit of the Marmot specification, completed in early 2026, is available from Least Authority at `https://leastauthority.com/`. On the NIP-17 gift-wrap envelope construction that Marmot uses for metadata-hiding message delivery, see `https://github.com/nostr-protocol/nips/blob/master/17.md`.

[34]For readers who want to use the tools covered in this chapter: Tor Browser (torproject.org) is the reference anonymity client; Mullvad (mullvad.net) and IVPN (ivpn.net) are the privacy-oriented VPN providers with the strongest public track records; Signal (signal.org) and Session (getsession.org) are the mainstream encrypted messengers; Element (element.io) is the reference Matrix client, and Wire (wire.com) ships MLS in production; Nym (nym.com) is the most developed deployed mixnet. For background reading, Matt Blaze's *Tangled Web* writings, the Tor Project's ongoing research notes, and the EFF's Surveillance Self-Defense guide (ssd.eff.org) are all worth consulting; Moxie Marlinspike, "The Ecosystem Is Moving" (2016), `https://signal.org/blog/the-ecosystem-is-moving/`, is the canonical modern statement of why federated protocols have struggled against integrated end-to-end-encrypted platforms and why the gap between deployed cryptography and academic cryptography keeps widening. Bruce Schneier, *Data and Goliath* (2015) and Carissa Véliz, *Privacy Is Power* (2020) address the wider political context around the same infrastructure.

18

BITCOIN AND THE DIGITAL MONEY BREAKTHROUGH

> "A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution."
>
> Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008)[1]

Digital money faced two fundamental problems that decades of cypherpunk research could not solve at once. The double-spending problem seemed to require a central authority, while the base-money problem meant most digital currencies were only claims on issuers, not money itself.

Satoshi Nakamoto's breakthrough solved both. This chapter traces the twin problems and the failed precursors that supplied partial answers. It ends with the consensus architecture that finally made decentralized digital money possible.

18.1 The Twin Problems of Digital Money

The Double-Spending Problem

Physical cash cannot be spent twice. When Alice hands Bob a gold coin, she no longer possesses it. The physical transfer is the

transaction. Physical objects are inherently scarce; possession by one excludes possession by another.

Digital information lacks this property. As Chapter 6 established, information is non-rivalrous: copying a file does not remove the original. If digital cash were a file that encoded value, Alice could send Bob a copy while keeping the original, then send the same "cash" to Carol. The double-spending problem is a scarcity problem: how can digital units be made rivalrous when digital information is inherently copyable?

Preventing double-spending requires establishing which transactions came first, thereby determining who possesses each unit. A database can record this, but who controls the database? Whoever controls the database can censor transactions or seize funds, and authorities can compel the operator to do either. The problem seemed to require centralization.

The Base Money Problem

Even if double-spending could be solved, a second problem remained: what exactly is being transferred?

Previous digital currencies were money substitutes: claims against issuers, not money proper. Users held account balances that companies promised to honor, not the money good itself: DigiCash balances were claims on DigiCash Inc., and E-gold balances were claims on e-gold Ltd. These systems created digital IOUs, not digital money.

Money substitutes require trust in the issuer. Issuers can fail, be shut down, refuse redemption, or be compelled by authorities to freeze accounts. A money substitute in your hand is a promise; money proper in your hand is the present good itself, valued for its monetary services without reference to any redeeming party. Physical gold requires no backing because the metal is what holders demand; a paper note promising gold requires trust that someone will honor the promise when presented with the note.

Creating digital money proper requires more than digital promises. A system must prevent double-spending without central

control, and it must constitute base money itself, not claims on an issuer. No system before Bitcoin achieved both.

18.2 Precursors and Their Failures

Bitcoin did not emerge from nothing; it combined earlier cypherpunk innovations and pushed past their limits.

DigiCash: Privacy and Double-Spending, Without Decentralization

David Chaum's DigiCash solved two problems at once. Blind signatures allowed the issuing bank to detect and reject duplicate tokens without being able to link withdrawals to deposits. Users could make untraceable payments, and the bank could prevent double-spending. Both problems seemed to require contradictory information flows, yet Chaum's cryptography threaded the needle.[2]

DigiCash failed on different grounds. It required a central server, creating a single point of failure that authorities could target. And it issued money substitutes: account balances were claims on DigiCash Inc., not money proper. When the company filed for bankruptcy in 1998, the system died with it. Chaum solved the cryptographic problem but not the institutional one.

E-gold: Backing Without Resistance

E-gold (1996-2009) attempted to create digital gold by maintaining physical gold reserves backing account balances. It attracted millions of users but remained a money substitute: account balances were claims on e-gold Ltd., not gold itself. The centralized structure made it vulnerable; US authorities shut down the operation and prosecuted its founders. E-gold showed that even commodity backing cannot substitute for decentralization.

Hashcash: Proof-of-Work as Access Control

Adam Back's Hashcash (1997) introduced proof-of-work: requiring computational effort to produce a token.[3] Originally designed to prevent email spam, Hashcash tokens shared one property with

base money: they required no issuer's promise or backing. Validity came from proof-of-work alone.

Hashcash, however, was never intended as money. It was an access control system in which tokens were bound to specific recipients by embedding the recipient's email address in the hashed data, and each mail server maintained its own database of tokens already seen, rejecting duplicates. That design prevented reuse of tokens at a single server but created no global scarcity: a token spent at one server had no effect on any other server, and there was no shared ledger or network-wide state that could give a token weight beyond its local context.

Money requires global consensus: all participants must agree on which units exist and who owns them. Hashcash had only local verification. It showed that computational work could create unforgeable tokens, but access rights are not money. The missing element was a shared, agreed-upon record of token ownership across all participants.

B-money and Bit Gold: The Inflation Problem

Wei Dai's B-money (1998) and Nick Szabo's Bit Gold (1998-2005) both proposed systems where proof-of-work directly created monetary units.[4] In B-money, the value of one unit was meant to equal the computational cost of producing it. In Bit Gold, valid hashes were the monetary units themselves.

Both suffered from a fundamental flaw: as computing power increases, tokens become cheaper to produce. Szabo recognized this explicitly: if it became possible to be "a low-cost producer (by several orders of magnitude)," one could "swamp the market with bit gold." His proposed solution was timestamping, so markets could value older hashes (harder to produce at the time) more than newer ones. But this destroys fungibility; not all units would be equal.

Neither solved the distributed consensus problem, and both remained theoretical. But their deepest issue was architectural: conflating proof-of-work with money creation. If work creates money, monetary policy depends on hardware economics.

The Missing Pieces

Each precursor solved part of the puzzle without closing the loop. DigiCash achieved privacy and solved double-spending but required centralization and issued money substitutes; e-gold achieved commodity backing yet remained a custodial claim; Hashcash showed unforgeable tokens through computational work, though as an access control system, not as money. B-money and Bit Gold articulated the architecture for decentralized digital money.

Two problems remained unsolved. First, none achieved decentralized consensus on a shared transaction history, the mechanism required to make digital units rivalrous without central control. Second, systems that used proof-of-work for money creation tied monetary policy to hardware economics, creating perpetual inflation as computing power grew.

Nakamoto solved both. His breakthrough joined decentralized consensus on transaction ordering to a full separation of proof-of-work from money creation.

18.3 Nakamoto Consensus

The Architecture

The problems are now clear. Double-spending requires global consensus on transaction ordering. Previous systems achieved this through central servers, which created single points of failure. Proof-of-work can create unforgeable tokens, but using it for money creation ties monetary policy to hardware economics. The challenge is to build a system where anyone can participate in ordering transactions, with no central authority to target, while monetary policy remains fixed regardless of computational growth.

Bitcoin's architecture combines a distributed ledger that anyone can read with consensus rules that every participant enforces independently, linked by a permissionless block production mechanism. The ledger records transaction history, block production extends it, and consensus rules determine what extensions are valid.

Anyone can propose a block of transactions to append to the ledger. No permission is required; no single entity controls what

gets recorded. But permissionless participation creates a problem: what prevents the system from being overwhelmed?

The Denial of Service Problem

If anyone can produce blocks without restriction, an attacker could flood the network. Each block must be downloaded and validated, then stored by every node. Even without malicious intent, if block production were free, rational participants would produce blocks constantly to collect fees and rewards. The network would drown in data.

The problem is larger than bandwidth alone, because verification also requires computational resources. Every transaction in every block must be checked for valid signatures, unspent inputs, and correct amounts. If blocks arrive faster than nodes can verify them, nodes fall behind, and if ordinary nodes cannot keep up, only those with exceptional resources can participate in verification. The system would centralize around whoever could afford the infrastructure, defeating its purpose.

A centralized system solves this trivially: the operator decides how many transactions to process. A decentralized system has no operator. The solution requires a throttling mechanism that emerges from the protocol itself, slowing block production without any authority deciding who may produce blocks or how often.

The target is roughly one block every ten minutes across the entire network, regardless of how many participants attempt to produce blocks. This rate is slow enough that ordinary hardware can verify all transactions, yet fast enough for practical use. The question is how to enforce this rate without a rate-limiter.

Proof-of-Work as Throttling

Proof-of-work provides the throttling mechanism. The insight comes from Hashcash: computational work that is difficult to produce but trivial to verify. Producing a valid block requires finding a specific kind of hash, which demands sustained computation.

Verifying that someone found it requires a single hash operation, nearly instantaneous.

A block header contains metadata: the hash of the previous block, a hash of included transactions, a timestamp, and a nonce (a variable field the miner can change freely). To produce a valid block, a miner must find a nonce such that hashing the entire header produces a number below a difficulty threshold.

Cryptographic hash functions like SHA-256[5] produce output that is effectively random given the input. Changing even one bit of input produces an entirely different hash. No mathematical relationship exists between input and output that would allow predicting which inputs yield low hashes. The only way to find a suitable nonce is repeated trial: set a nonce, compute the hash, check if it meets the threshold, increment the nonce, repeat. Miners perform billions of these operations per second.

This process provably takes time. The difficulty threshold determines how many attempts are needed on average. If the threshold requires a hash starting with 20 zero bits, approximately one in a million hashes will qualify. Requiring 30 zero bits means approximately one in a billion. The work cannot be faked or shortcut. Verification checks whether the hash meets the threshold, and anyone can verify instantaneously by computing a single hash.

The asymmetry is essential. Production is expensive; verification is cheap. This allows any node to validate blocks without trusting the miner, while making block production costly enough to throttle the rate.

Difficulty Adjustment

Proof-of-work throttles block production, but the throttle must adapt. If difficulty were fixed, increasing computational power would produce blocks faster, overwhelming the network. Decreasing power would slow blocks to a crawl, making the system unusable. The system needs a feedback mechanism.

Every 2016 blocks (approximately two weeks at the target rate), the protocol recalculates difficulty. It compares the actual time elapsed since the previous adjustment to the expected time (2016

blocks times ten minutes). If blocks arrived faster than ten minutes on average, difficulty increases; the threshold lowers, requiring hashes with more leading zeros. If blocks arrived slower, difficulty decreases; the threshold rises, accepting hashes that would previously have been rejected.

The adjustment is bounded: difficulty cannot change by more than a factor of four in either direction per period. This prevents extreme oscillations if hash rate changes dramatically. The bounds also limit potential manipulation; miners cannot game timestamps to radically lower difficulty.

This mechanism makes Bitcoin self-regulating. When mining becomes more profitable, for example because the price rises, more computational power enters the network; blocks would arrive faster, but difficulty adjusts upward and restores the ten-minute average. When profitability falls, miners exit, and difficulty adjusts downward to compensate. The system finds equilibrium at whatever level of computational power the market provides.

The result is that increased hash rate produces more security, not more blocks and not more bitcoin. A network with ten times the hash rate is ten times more expensive to attack, but still produces blocks at the same rate with the same reward schedule. This decouples security from monetary policy, a property no predecessor achieved.

What Proof-of-Work Does Not Do

Proof-of-work does not create bitcoin. This distinction separates Bitcoin from its predecessors.

In B-money and Bit Gold, proof-of-work was supposed to create monetary units directly: the work itself was the money. This tied monetary policy to hardware economics. As computing power grew cheaper, money creation would accelerate indefinitely.

Bitcoin inverts this relationship. Proof-of-work throttles block production and orders transactions; it does not determine how much bitcoin exists. The block reward is defined by consensus rules that every full node validates, not by the amount of work performed. A miner who claims a larger reward produces an invalid block,

regardless of how much work went into it. Difficulty adjustment ensures that increased computing power produces more security, not more bitcoin.

Monetary policy in Bitcoin is enforced by full nodes run by economically relevant participants, not by miners performing work. Miners propose blocks; nodes accept or reject them according to fixed rules. This separation is why Bitcoin's supply schedule remains unchanged despite hash rate increasing by orders of magnitude since launch.

Fair Issuance Through Mining

If proof-of-work does not create bitcoin, how do new coins enter circulation? The answer reveals another role for mining: distribution mechanism.

Each valid block includes a coinbase transaction that creates new bitcoin according to a consensus-defined schedule. The initial reward was 50 BTC per block. Every 210,000 blocks (approximately four years), the reward halves: 25, then 12.5, then 6.25, then the current 3.125 BTC. This halving continues until approximately 2140, when the last fraction of a bitcoin is mined and the total supply reaches 21 million.[6]

The miner who produces a valid block receives this reward. Since block production requires proof-of-work, and since anyone can attempt to mine, issuance operates as a continuous open lottery. Every ten minutes on average, the network awards new bitcoin to whoever finds the next valid block.

This mechanism has properties most later monetary systems did not achieve. New coins go to those who expend real resources securing the network, not to insiders, political favorites, or early investors. No premine allocated coins to founders before the network launched, and no ICO sold tokens to speculators. No central authority decides who receives new issuance. The first block Satoshi mined followed the same rules as every subsequent block.

The contrast with alternatives is stark. Fiat currencies are issued by central banks to governments and politically connected institutions. Proof-of-stake systems reward existing holders proportionally

to their holdings; the rich get richer by definition. Many cryptocurrencies launched with premines or insider allocations that enriched founders before public participation was possible.

Bitcoin's mining lottery ensures that anyone willing to expend resources can compete for new issuance. Geographic location does not matter. Political connections do not matter. Existing wealth provides no special claim. A miner in any jurisdiction, operating any scale of equipment, has a probability of winning proportional to their share of network hash rate. The playing field is not perfectly level, as industrial miners have economies of scale, but it is open. No permission is required to participate.

This transforms issuance from a political question into a market process. New bitcoin flows to those who provide the service the network needs: computational security. The work is not wasted; it throttles block production and orders transactions while distributing new coins to those who protect the ledger.

Chain Selection and Consensus

Bitcoin maintains a blockchain: an ordered sequence of blocks, each referencing the hash of its predecessor. The first block (the genesis block)[7] has no predecessor; every subsequent block commits to the entire history before it. Changing any transaction in any historical block would change that block's hash, which would invalidate the reference in the next block, which would change its hash, cascading forward to the present. The structure makes history tamper-evident.

But tamper-evidence is not consensus. Multiple valid chains could exist. When a miner finds a valid block, they broadcast it to the network. Other miners, upon receiving it, face a choice: build on this block, or continue working on their current candidate. Network latency means different miners see different blocks at different times. Two miners might find valid blocks at nearly the same moment, each unaware of the other. The network temporarily has two valid chain tips.

The protocol treats the temporary fork as expected operation and specifies a resolution rule: nodes follow the chain with the most ac-

cumulated proof-of-work, so when one branch gets extended before the other, the extended branch has more work, and miners abandon the shorter branch to build on the longer one. The abandoned block becomes orphaned; its transactions return to the mempool for inclusion in future blocks.

The rule means that transactions become more secure over time. A transaction in the most recent block could be displaced if that block is orphaned. A transaction buried under six blocks would require an attacker to produce a competing chain with more work than six blocks' worth, an increasingly expensive proposition. Deep transactions are practically irreversible.

Nakamoto's original analysis goes further. The catch-up attempt can be modeled as a probability problem, not only an economic one. In the Bitcoin whitepaper's Section 11, an attacker who tries to rewrite recent history is treated as a gambler playing against the house: each new block is a contest in which the honest network either extends its lead or gives the attacker a chance to close it. The whitepaper's only non-computer-science reference, William Feller's *An Introduction to Probability Theory and Its Applications* (1957), supplies the Gambler's Ruin mathematics behind this framing.[8] If the attacker commands less than half the network's hash power, the probability of ever catching up falls exponentially with the number of confirmations the honest chain has already accumulated. Six confirmations, the common rule of thumb, flows directly from this analysis. Bitcoin's security against deep reorganization is therefore emergent: probabilistic dynamics make the attack overwhelmingly unlikely to succeed, and the improbability compounds with each additional block.

This mechanism achieves consensus through accumulated work, with no voting and no predetermined validator set. The chain with most accumulated work is network agreement on transaction order. Nakamoto called it a "proof-of-work chain" that serves as "proof of the sequence of events witnessed." Miners do not vote; they produce blocks. Nodes independently select the chain with most accumulated proof-of-work.

Why Miners Validate

Miners receive block rewards (newly created bitcoin) and transaction fees for producing valid blocks. But why do miners bother validating transactions and following protocol rules? They could produce blocks containing invalid transactions or claiming excess rewards. Nothing in proof-of-work itself prevents this; the hash function does not know whether the block contents are valid.

The answer lies in the economic structure, specifically in who defines what counts as bitcoin.[9]

Economically relevant participants such as merchants and exchanges often run full nodes. A full node independently validates every transaction against protocol rules: correct signatures, unspent inputs, valid amounts, proper block reward. A node does not ask anyone whether a transaction is valid; it checks for itself. If a block violates any rule, the node rejects it. The proof-of-work is irrelevant; invalid blocks are discarded regardless of how much computation went into producing them.

Miners produce blocks, but the validating economy defines the valid chain. A miner who produces an invalid block, whether claiming an inflated reward, including a double-spend, or violating any consensus rule, has wasted their computational resources. Economically relevant nodes will reject that block. The block reward exists only on a chain that the wider economy does not recognize. It cannot circulate as bitcoin if counterparties refuse it.

This creates the enforcement mechanism. Miners validate because the wider economy validates. The block reward and transaction fees are worthless unless counterparties accept them as payment. Since economically relevant nodes accept bitcoin only from the valid chain with most accumulated work, miners are economically compelled to produce valid blocks on that chain.

The relationship is subtle but essential. Miners do not control Bitcoin; they serve it. A miner with 51% of hash power could theoretically reorganize recent history or censor transactions, but cannot change the rules. Attempting to overclaim the block reward or spend coins they do not own would produce an invalid chain that merchants reject. Hash power without validity is worthless.

This structure inverts common intuitions about how the system is secured. Security does not come from miners being trustworthy; it comes from independent validation, because every economically relevant node is an enforcement point and the more participants validate independently, the more resilient the system becomes. Those who do not validate, who trust someone else's node, have delegated their enforcement power and depend on that delegate's honesty.

Bitcoin as Commitment Device

Thomas Schelling's 1960 analysis of strategic interaction introduced the commitment device as a tool that rational parties use to constrain their own future actions in ways that make threats and promises credible.[10] A government that wants a low-inflation reputation benefits from a central bank whose institutional design makes high inflation costly for its governors. A contract that binds both parties is worth more than a contract either side can walk away from. A protocol whose rules cannot be changed by the operator is more useful as a commitment than a protocol whose rules can be rewritten.

Bitcoin's architecture is a commitment device in this sense. The 21-million-coin supply cap is a number in the consensus rules, and a sufficiently coordinated majority of the economy could change it; the economy that would change it is the same economy whose savings the cap protects, which is why the cap has held across sixteen years of attempts to renegotiate it. The difficulty adjustment binds hashrate growth to block timing, so that increased resources committed to the chain's security cannot accelerate issuance or rewrite the monetary schedule. The proof-of-work chain is a commitment to whichever transaction ordering accumulated the most work, which binds every participant to treat the chain they join as the chain they will extend. Each is a self-imposed constraint participants honor because the value of honoring it depends on everyone else honoring it too.

Nick Szabo's earlier work on smart contracts and the God protocols brought the same framing into the cypherpunk tradition.[11] A contract executed by code is a commitment device whose enforce-

ment is the code's determinism, and participants extend trust to the deterministic execution instead of to any enforcer's discretion. Bitcoin works as money because the protocol's commitment is encoded in data every participant validates, and every participant's validation is independent of any other participant's willingness to honor a promise.

Chapter Summary

Digital money failed for years because two problems remained unsolved at once. Systems could protect privacy without decentralization or decentralize some functions without producing money proper, but none could prevent double-spending without reintroducing trusted control. Bitcoin succeeded by combining previously separate ideas into one architecture.[12] Proof-of-work throttled denial-of-service and made cost imposition measurable. Chain selection solved transaction ordering without a central ledger: the chain with the most accumulated work is the chain, and an attacker who tries to rewrite recent history is playing a Gambler's Ruin whose probability of success falls exponentially with each additional confirmation. Mining issues the asset through the same process that defends the network, which separates security from monetary policy, increased hash rate produces more security, not more bitcoin. This distinguishes Bitcoin from B-money and Bit Gold, where proof-of-work created monetary units directly and tied supply to hardware economics.

Bitcoin is base money. Unlike DigiCash balances or e-gold accounts, bitcoin units are not claims on an issuer. Holders possess the units themselves, verified by running a full node against consensus rules that no single entity controls. Miners produce blocks, but the validating economy defines the valid chain: a miner who tries to overclaim the block reward or spend coins they do not own produces an invalid chain that merchants and exchanges reject, and hash power without validity is worthless. The architecture is a commitment device in Schelling's sense, rules that every participant can verify independently bind everyone to the same ledger, and no

single authority can alter them without convincing the network to run different software.

Bitcoin's monetary properties, its resistance to shutdown, and the empirical survival record belong to Chapter 19. Its privacy model, the base layer's transparency, chain-analysis threats, and the layered privacy tools built around it, belongs to Chapter 20. The chapter argues that proof-of-work solved specific problems its predecessors could not; alternative consensus mechanisms exist and their comparison is the subject of separate work.

Notes

[1]Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008), `https://bitcoin.org/bitcoin.pdf`. First introduced at Chapter 2, note 8, which covers Bitcoin's canonical project pages and launch context.

[2]David Chaum, "Blind Signatures for Untraceable Payments," *Advances in Cryptology: Proceedings of Crypto 82*, ed. David Chaum, Ronald L. Rivest, and Alan T. Sherman (Plenum Press, 1983), 199–203, is the foundational paper introducing blind signatures. Chaum, "Security Without Identification: Transaction Systems to Make Big Brother Obsolete," *Communications of the ACM* 28, no. 10 (1985): 1030–1044, extends the argument to a whole-system design. For Hal Finney's 1993 explanation of how Chaumian ecash detects double-spending while preserving anonymity until the second spend, see Finney, "Detecting Double Spending," `https://nakamotoinstitute.org/library/detecting-double-spending/`. DigiCash's company history is documented in Steven Levy, *Crypto* (Viking, 2001), chapter 8; on DigiCash's bankruptcy (1998), see Julian Dibbell, "In Gold We Trust," *Wired* (January 2002). On e-gold's shutdown: the U.S. Department of Justice charged the operators of e-gold Ltd. in April 2007 under 18 U.S.C. §1960 (operating an unlicensed money-transmitting business); the principals pleaded guilty in 2008 and the service ceased redemptions in 2009. *United States v. e-gold, Ltd.*, No. 1:07-cr-00109 (D.D.C. 2007).

[3]Adam Back, "Hashcash: A Denial of Service Counter-Measure," Technical Report (2002), `http://www.hashcash.org/papers/hashcash.pdf`. The original proposal was announced on the cypherpunks mailing list in March 1997. For the bridge from non-reusable Hashcash stamps to transferable tokens backed by trusted hardware, see Hal Finney's RPOW announcement (2004), `https://nakamotoinstitute.org/library/rpow/`. For the broader proof-of-work literature that Hashcash descends from, Cynthia Dwork and Moni Naor, "Pricing via Processing or Combatting Junk Mail," *CRYPTO '92* (1993), 139–147, introduced the idea of using computational cost as an access-control mechanism.

[4]Wei Dai, "b-money," unpublished proposal (November 1998), http://www.weidai.com/bmoney.txt. Nick Szabo, "Bit Gold" (1998, published on Szabo's blog in 2005), https://unenumerated.blogspot.com/2005/12/bit-gold.html. Szabo's later "Bit Gold Markets" (2008), https://nakamotoinstitute.org/library/bit-gold-markets/, makes explicit the non-fungibility problem this chapter highlights: because puzzle solutions from different periods would have different production costs, markets would have to pool and tranche them into standard-value bundles. For the timestamping lineage that Bit Gold presupposes, see W. Scott Stornetta and Stuart Haber, "How to Time-Stamp a Digital Document" (1991), https://nakamotoinstitute.org/library/time-stamp-digital-document/. Satoshi Nakamoto cited Dai's b-money in the Bitcoin whitepaper; the relationship between Bit Gold and Bitcoin is more elliptical but the conceptual debt is clear.

[5]SHA-256 (Secure Hash Algorithm 256-bit) was designed by the U.S. National Security Agency and published by NIST in 2001 as FIPS PUB 180-2. Bitcoin uses a double application of SHA-256 (SHA-256d) for block hashing and a SHA-256/RIPEMD-160 combination for address derivation. The algorithm specification is available at National Institute of Standards and Technology, "Secure Hash Standard (SHS)," FIPS PUB 180-4 (2015), https://csrc.nist.gov/publications/detail/fips/180/4/final. For the security properties that make SHA-256 suitable for proof-of-work — collision resistance, preimage resistance, and the avalanche effect — see Alfred J. Menezes, Paul C. van Oorschot, and Scott A. Vanstone, *Handbook of Applied Cryptography* (CRC Press, 1996), chapter 9, freely available at https://cacr.uwaterloo.ca/hac/.

[6]The Bitcoin supply schedule and halving mechanism are specified in the Bitcoin whitepaper, section 6 ("Incentive"), and encoded in Bitcoin Core's `GetBlockSubsidy()` function. The 210,000-block halving interval and 21-million-coin cap are consensus rules enforced by every full node. For the on-chain record of each halving event, see https://mempool.space/mining. The economic implications of the fixed supply and disinflationary issuance are analyzed in Saifedean Ammous, *The Bitcoin Standard: The Decentralized Alternative to Central Banking* (Wiley, 2018), chapters 1–3, which situates Bitcoin's supply schedule against the stock-to-flow dynamics of gold and silver. For a critical examination of the fee-transition question (what happens to miner revenue when block subsidies approach zero), see Miles Carlsten et al., "On the Instability of Bitcoin Without the Block Reward," *ACM CCS* (2016).

[7]The genesis block (block 0) was mined by Satoshi Nakamoto on January 3, 2009. Its coinbase transaction contains the text "The Times 03/Jan/2009 Chancellor on brink of second bailout for banks," a reference to that day's front-page headline in *The Times* (London), functioning as a timestamp and a statement of purpose. The block hash is 000000000019d6689c085ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f. It is viewable at https://mempool.space/block/000000000019d6689c08

5ae165831e934ff763ae46a2a6c172b3f1b60a8ce26f. Uniquely, the 50 BTC coinbase reward in the genesis block is unspendable: it was not included in the initial unspent transaction output set. For a detailed analysis of the genesis block's technical properties and embedded message, see Jameson Lopp, "Bitcoin's Genesis Block," https://blog.lopp.net/bitcoins-genesis-block/.

[8]William Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1 (New York: John Wiley & Sons, 1957), is cited in the Bitcoin whitepaper as reference [8] and supplies the Gambler's Ruin analysis that Nakamoto uses to bound the probability of an attacker catching up to the honest chain. Bitcoin's resistance to deep reorganization is not a wall the protocol builds around transactions but an emergent property of probabilistic dynamics under adversarial conditions, with the probability of successful attack falling exponentially as confirmations accumulate. The whitepaper's Section 11 contains the full derivation.

[9]On the economic relationship between merchants and miners, see Eric Voskuil, "Qualitative Security Model," *Cryptoeconomics* (2020), https://github.com/libbitcoin/libbitcoin-system/wiki/Qualitative-Security-Model. Voskuil emphasizes that individuals who do not personally validate have delegated validation to a central authority, undermining the security model. For the wider treatment of Bitcoin's security model in this book, see Chapter 5, note 9.

[10]Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, MA: Harvard University Press, 1960), especially chapter 2 ("An Essay on Bargaining"). Schelling's later work in *Arms and Influence* (Yale University Press, 1966) develops the commitment-device analysis for strategic interaction more generally. For the economic application to monetary institutions, Finn E. Kydland and Edward C. Prescott, "Rules Over Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy* 85, no. 3 (1977): 473–491, formalizes the time-inconsistency problem that central-bank commitment devices address.

[11]Nick Szabo, "Smart Contracts: Building Blocks for Digital Markets" (1996), and "The God Protocols" (1997), both at https://nakamotoinstitute.org/the-god-protocols/ and https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html. Szabo's "Shelling Out: The Origins of Money" (2002), https://nakamotoinstitute.org/shelling-out/, develops the framing across commodity-money history. For the cryptoeconomic lineage connecting commitment devices to consensus protocols, see Vitalik Buterin, "A Next-Generation Smart Contract and Decentralized Application Platform" (2014), and Eric Voskuil, "Cryptoeconomics" (2020) already cited in note 6. For the earliest Bitcoin-era primary sources, Hal Finney's posts to the cypherpunks mailing list (1992–2004) and to Bitcointalk (2009–2013) are archived at https://bitcointalk.org/index.php?action=profile;u=2435 and, for the earlier cypherpunk-era correspondence, at https://cypherpunks.venona.com; Finney received the first Bitcoin

transaction from Satoshi on January 12, 2009 and posted one of the first public analyses of Bitcoin's privacy properties. For a historical treatment of these precursors and their relationship to Bitcoin, see Finn Brunton, *Digital Cash: The Unknown History of the Anarchists, Utopians, and Technologists Who Created Cryptocurrency* (Princeton University Press, 2019), and Nathaniel Popper, *Digital Gold* (HarperCollins, 2015), chapters 1–3. On Byzantine fault tolerance, the classical underlying problem Bitcoin's consensus solves probabilistically, see Leslie Lamport, Robert Shostak, and Marshall Pease, "The Byzantine Generals Problem," *ACM Transactions on Programming Languages and Systems* 4, no. 3 (1982): 382–401, `https://lamport.azurewebsites.net/pubs/byz.pdf`. For the comparative lineage across all these precursors, see Arvind Narayanan and Jeremy Clark, "Bitcoin's Academic Pedigree," *Communications of the ACM* 60, no. 12 (2017): 36–45.

[12]For a textbook-level synthesis of how Bitcoin combines proof-of-work, the blockchain data structure, and distributed consensus, see Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder, *Bitcoin and Cryptocurrency Technologies* (Princeton University Press, 2016), freely available at `https://bitcoinbook.cs.princeton.edu/`. Chapter 2 covers the cryptographic primitives (hash functions, hash pointers, digital signatures) that underpin the architecture described here, and chapter 3 covers the mechanics of Bitcoin's consensus protocol. Narayanan and Clark's companion article, "Bitcoin's Academic Pedigree," *Communications of the ACM* 60, no. 12 (2017): 36–45, traces how each component of Bitcoin's design draws on prior academic work, providing the most complete map of Bitcoin's intellectual lineage available in a single source.

19

BITCOIN AS SOUND AND RESISTANT MONEY

> "I don't believe we shall ever have a good money again before we take the thing out of the hands of government, that is, we can't take it violently out of the hands of government, all we can do is by some sly roundabout way introduce something that they can't stop."
>
> F. A. Hayek, *Denationalisation of Money* (1976)[1]

Chapter 18 showed how Bitcoin becomes possible. This chapter examines the monetary properties encoded in Bitcoin and the resistance properties that let the system survive opposition.

These two dimensions belong together. A digital asset that can be inflated, frozen, or easily shut down is not sound money in any serious sense. Bitcoin builds monetary discipline and resistance into the same architecture.[2]

19.1 Sound Money Properties as Schelling Points

Chapter 9 established the properties that sound money requires. Bitcoin exhibits these properties, and the mechanism that upholds them is worth getting right at the outset, because every claim in this section depends on it.

A consensus rule in Bitcoin is not enforced by the protocol in the way a contract is enforced by a court. There is no authoritative

copy of the ruleset. Every full node runs its own software, holds its own consensus rules, and validates every block against those rules independently. Any operator can modify their own rules, run a fork that accepts blocks the rest of the network rejects, and produce an alternate chain on that basis. Nobody can compel another operator to do the same. The rules persist because independent validators have converged on them and refuse to recognize chains that violate them. The cap, the issuance schedule, the signature rules, the block subsidy, and every other consensus rule discussed below holds the same way: as a Schelling point on which validators have coordinated, where unilateral deviation produces a chain with no economic value because no other validator will accept it.

The convergence model predicts the actual behavior of the system in cases where the alternative framing fails. Bitcoin Cash forked off Bitcoin in 2017 by changing the block-size rule, and the resulting chain has its own market and is irrelevant to Bitcoin holders because Bitcoin validators did not converge on the new rule. Ethereum rolled back the DAO transactions in 2016 by social agreement among Ethereum validators, and the rollback succeeded because validators converged on it; the minority that refused now runs Ethereum Classic.[3] In both cases the rule that prevailed was upheld by convergence among independent operators, and every property in this section holds the same way.

Fixed Supply Cap

Bitcoin's supply is capped at 21 million coins. The cap is a consensus rule that every full node validates against blocks it receives, and a block containing transactions that create coins beyond the schedule is rejected as invalid by every node running rules consistent with that cap. A miner who produces such a block can publish it, but the chain extending from it is recognized only by nodes running modified rules, and no exchange, custodian, or merchant running standard rules will accept payments on it. The block has no economic value, so no rational miner produces one.

The cap is therefore immutable in the practical sense that holders care about: there is no path to changing the rule on the chain you

are running short of you yourself running different software. A miner or developer who wants more supply can fork off and build a chain with a higher cap, but they cannot pull the original chain along with them. The original chain keeps running for everyone who keeps running the original rules, even if that is a minority of former participants. Unlike physical gold, where new deposits can be discovered, or fiat currency, where central banks can print, there is no entity with the power to increase the supply on the chain that holders are validating; an attempt to do so produces a separate chain that those holders ignore. Because solving double-spending makes bitcoin units rivalrous (one person's possession excludes another's), property rights can apply to this digital asset, and the cap on the asset class is itself a property right held in common by every node operator who refuses to accept inflation.

Predictable Issuance

The difficulty adjustment and the halving schedule means the entire supply curve is known in advance. Anyone can calculate future supply at any date. There is no emergency decree or discretionary policy that could change it, because there is no authority empowered to issue one; changing the schedule requires the same coordination problem as changing the cap, and validators who have refused inflation for seventeen years have no incentive to accept it now. Monetary policy is transparent because it is the same rule on every operator's machine, and it is durable because deviation from it produces a chain nobody values.

Divisibility

Bitcoin is divisible to eight decimal places. The smallest unit, one hundred-millionth of a bitcoin, is called a satoshi. This enables transactions of any practical size, from micropayments to large settlements. The eight-decimal design was not Satoshi's first sketch. Early correspondence with Hal Finney in January 2009 pushed for finer divisibility than the original draft contemplated, and the 10^8

satoshi-per-coin parameter was settled in that exchange before the network went live to a wider audience.[4]

This divisibility is part of Bitcoin's current unit definition, not a physical limit. If future adoption requires finer subdivision, additional decimal places would require a consensus change. Divisibility is a parameter, not a fixed constraint.

No physical cutting or melting is required. A single bitcoin can be divided into 100 million satoshis without losing value or requiring trust in a third party. This exceeds the divisibility of any physical commodity.

Portability

Bitcoin can be transferred anywhere on earth with network connectivity. Value crosses borders without physical transport, customs inspection, or confiscation risk at checkpoints.

A private key, which controls any amount of bitcoin, can be memorized, stored on a device, or encoded in various formats. Carrying a billion dollars in gold requires trucks and guards; carrying a billion dollars in bitcoin requires remembering twelve words.

Durability

Bitcoin does not decay. Units created in 2009 remain identical to units created today. There are no storage costs for the asset itself, though maintaining secure access to private keys requires care.

The ledger is replicated across thousands of nodes worldwide. Individual storage media fail, but the network maintains redundant copies. This replication protects against data loss at the network level, though individual users can still lose access to their coins through lost keys, forgotten passwords, or destroyed backups. Gold and bitcoin share this vulnerability: both can be lost through carelessness or disaster. Bitcoin's durability advantage over physical currency is clearer; paper money degrades, coins corrode, but bitcoin units remain cryptographically intact indefinitely.

Verifiability

Anyone running a full node can verify the entire monetary history. Unlike gold, which requires assay, or banknotes, which require specialized equipment to detect counterfeits, bitcoin authenticity is cryptographically certain. A valid transaction either satisfies the protocol rules or it does not; no judgment or expertise is required beyond running the software.

This verification extends to supply. Any participant can independently confirm that total supply follows the schedule. No central authority's statement must be trusted. The ledger itself is the proof.

Fungibility Challenges

Sound money requires that units be interchangeable. A dollar is a dollar regardless of its history. Gold bars are fungible; one ounce equals any other ounce of the same purity.

Bitcoin's transparent blockchain creates fungibility challenges. Every transaction is recorded; every unit has a traceable history. Some exchanges and services reject coins with histories involving sanctioned addresses, darknet markets, or ransomware payments. If units are not interchangeable due to their history, fungibility is compromised.

Compromised fungibility is a real limitation of base layer Bitcoin. Later sections examine privacy tools such as CoinJoin,[5] PayJoin, Lightning,[6] and ecash[7] that address fungibility by breaking the links that enable discrimination.

How Validators Uphold Soundness

Traditional monetary soundness depends on institutional promises. Central banks promise stable policy and often break promises. Gold standards promise convertibility, and governments suspend convertibility.

Bitcoin's soundness is upheld by every participant who runs validating software and refuses chains that violate the rules they hold. Full nodes verify every transaction against the consensus rules each operator has chosen to run. Transactions that violate those

rules, including any that exceed supply limits, are rejected by the operator's node and never enter the chain that operator recognizes. Each node holds its rules unilaterally; nobody can compel a node to accept a chain that violates them. Coordinated change among many operators produces a fork: the original chain continues running for everyone still on the original rules, the modified chain runs alongside it for those who switched, and the question of which chain carries monetary value is settled by markets, exchanges, merchants, and liquidity, not by any threshold of validator participation. Inflation, censorship, and rollback have been resisted in Bitcoin's history because the holders who would lose under those changes have kept running the rules that protected them, and no one could take those rules away from them.

The system is not "trustless" in the sense of requiring no trust. Users trust that the software correctly implements the rules they want, and that the chain they are following will continue to attract the economic activity that gives their coins value. The first trust is verifiable: anyone can read the code and run a node that confirms each block obeys the rules they care about. The second is empirical: holders watch the markets, exchanges, and merchants that price and accept the chain they validate.

Bitcoin and the Regression Theorem

Chapter 9 presented the regression theorem and the questions it raises for novel moneys. Here we apply that framework directly to Bitcoin.

The problem appears simple: Mises showed that money's current value traces back through prior valuations to a time when the money commodity was valued for non-monetary use. Gold was ornament before it was money. But Bitcoin was designed as money from the start. Does it violate the theorem?

The resolution lies in the subjective theory of value itself. The theorem requires that first valuers had reasons for valuing; it does not specify what kinds of reasons qualify. Early Bitcoin adopters valued it for various reasons: ideological commitment to cypherpunk goals, technical fascination with the cryptographic innovation,

speculative anticipation of future adoption, or practical desire for censorship-resistant transactions. Each of these is a subjective valuation. Praxeology provides no basis for declaring some valuations legitimate and others illegitimate.

Bitcoin's original utility was real: it enabled permissionless, censorship-resistant transactions that no other system could provide. This utility is distinct from monetary use; one could value Bitcoin for this capability without expecting it to become generally accepted money. A market price for bitcoin in fiat existed before any goods trade: BitcoinMarket.com, the first BTC-USD exchange, opened in March 2010 and produced a continuous quoted price two months before the pizza event.

The May 22, 2010 transaction is the canonical illustration of how that price anchors into commerce. Laszlo Hanyecz posted on the BitcoinTalk forum offering 10,000 BTC for two pizzas delivered to his door, and a forum user (Jeremy Sturdivant, "jercos") accepted by ordering Papa John's pizzas with his own dollars. Papa John's never handled bitcoin; the BTC moved from Laszlo to Jeremy in exchange for the fiat-purchased pizzas. The praxeological content is in that peer-to-peer trade: Jeremy's willingness to accept 10,000 BTC for goods worth roughly twenty-five dollars and Laszlo's willingness to part with them at that rate are two subjective valuations meeting at a price already framed by exchange trading.

Direct merchant acceptance came later. By late 2010 the BitcoinTalk Marketplace board hosted vendors taking bitcoin directly for goods, with the alpaca-sock vendor Michael Stoneman emblematic of the period. Silk Road's launch in February 2011 made direct BTC commerce routine at scale, with buyers and vendors settling end-to-end in bitcoin without any fiat-rail intermediary. The regression chain runs through all of these.[8]

The theorem's core insight is that money emerges through market process, not decree. Bitcoin validates this insight more purely than any historical example. No legal tender law compelled acceptance, no government backing supported it, and no commodity convertibility anchored it. Market participants adopted bitcoin because they valued its properties, and that voluntary adoption

produced monetary status. The emergence shows that the theorem explains how money typically develops; it does not restrict which goods can become money.[9]

19.2 Resistance Properties: Why Bitcoin Survives

Decentralization Prevents Single-Point Shutdown

Bitcoin has no headquarters and no CEO. It also has no single server to seize. The network consists of thousands of nodes worldwide, each independently validating transactions. Eliminating Bitcoin would require shutting down all nodes simultaneously across every jurisdiction.

Previous digital currencies failed because authorities could target central points. Bitcoin's distributed architecture eliminates such targets.

Global Distribution

Bitcoin nodes operate in every major country. Mining occurs across continents. Development happens across jurisdictions. This geographic distribution means no single government controls the network.

Even concerted multinational action faces coordination problems, because different governments have different interests. While some jurisdictions restrict Bitcoin, others embrace it, and the network routes around restrictions.

Network-Level Attack Vectors

Decentralization does not eliminate all attack vectors. Network-level attacks can disrupt Bitcoin without targeting individual nodes.[10]

BGP hijacking allows autonomous systems to divert Bitcoin traffic by announcing false routing information. Research has shown that hijacking fewer than 900 IP prefixes could partition significant portions of the network. An ISP carrying Bitcoin traffic can delay block propagation by 20 minutes while remaining undetected. Such

attacks can cause chain splits and double-spending opportunities during the partition, along with loss of mining revenue when orphaned blocks are discarded after the attack ends.

Eclipse attacks target individual nodes by monopolizing their peer connections, isolating them from the honest network. The original eclipse attack research (2015) prompted improvements in Bitcoin Core's peer selection and connection handling. Current versions use several mitigations, including diverse outbound connections across different network groups, anchors that persist across restarts, and detection of suspicious peer behavior. The specific attacks described in early research are largely mitigated, though the attack class remains a concern that ongoing development continues to address.

DNS-based attacks can disrupt node discovery, and ISP-level blocking can impair operation in specific jurisdictions. China's 2021 mining ban[11] changed mining geography, showing that jurisdictional action can affect the network even without eliminating it entirely.

These attacks illustrate the distinction between implementation vulnerabilities and structural constraints. Implementation vulnerabilities (specific eclipse attack vectors, peer selection weaknesses) are addressed through ongoing Bitcoin Core development; the project maintains active security review and regularly releases updates. Structural constraints (reliance on internet routing infrastructure, BGP vulnerabilities) cannot be fully eliminated at the application layer, though mitigations like Tor usage and diverse connectivity help. Nodes can use Tor[12] to hide their IP addresses, reducing exposure to some network-level attacks.

Mining Concentration

A mining pool is a cooperative arrangement among many independent miners who combine their hash power to smooth their income. Solo mining at modern difficulty produces highly variable returns: a small operator might wait years between blocks, with long stretches of zero income punctuated by occasional full block rewards. Pools fix this by aggregating miners under a coordinator

who builds candidate block templates, distributes them to members, collects "shares" (low-difficulty proofs that members are hashing on the assigned templates), and pays out the eventual block reward in proportion to the shares each member contributed. The miner trades a small operator fee for steady income; the pool delivers payouts large enough to attract participation. The mining hardware stays with the individual miners; but the coordination layer and block template construction is centralized.

The pool layer has concentrated in practice. A small number of pools coordinate the majority of hash rate; as of recent measurements, the top five pools often control over 70% of it.[13]

The concrete vulnerability is template construction. Because the pool builds the candidate block, the pool chooses which transactions appear in it and which prior block the template extends. An operator under legal or political pressure can refuse to include particular transactions, and the miners hashing for that pool produce the censoring blocks without knowing the difference.

The same mechanism enables reorg attempts. A pool that points its templates at an older block instead of the current tip directs its members' hash power into an alternative chain that, if it wins the race against the honest tip, replaces the recent history with a different one. Whether the reorg succeeds depends on hash rate; a majority pool can force shallow reorgs at will, and a substantial-minority pool can occasionally succeed against the most recent blocks before they accumulate confirmations.

The hard limit is theft. A pool cannot spend coins whose private keys it does not hold, no matter how much hash rate it commands. Reorg attacks can double-spend the attacker's own recently-broadcast transactions, but they cannot take other holders' funds. Individual miners can switch pools at low cost, and pools that censor visibly or attempt visible reorgs would lose them to competitors, which is the disciplining mechanism in practice.

Stratum V2 and Datum address the architectural problem directly. Both protocols let individual miners or mining farms construct their own block templates and submit them to the pool, with the pool retaining only the share-aggregation and payout role. A

miner using either remains in the pool's payout system but picks which tip to extend and which transactions to include, decoupling template censorship and reorg-direction from coordination scale.[14]

Why Proof-of-Work Admits Perpetual Competition

The resistance properties above depend on a specific design choice: Bitcoin ties block production to the consumption of a resource produced outside the ledger. A miner votes for a chain by spending electricity on specialized hardware that was manufactured, shipped, and installed in the physical world. The security resource, hashrate, is a higher-order good assembled from silicon, energy, capital, and labor that exist independently of Bitcoin itself. This has a consequence most discussions of consensus understate: the supply of hashrate is exogenous to the protocol. No attacker can corner it.

Proof-of-stake systems work differently. In a stake-based protocol, the resource that grants voting rights is the coin itself. An attacker who captures a majority of active stake does not face a minority that can respond by manufacturing more of the security resource. The minority can only acquire more coins by purchasing them, and in a proof-of-stake system the attacker is the largest coinholder by definition. The attacker need not sell, and the remaining holders do not collectively control a majority. Slashing can destroy misbehaving stake when a violation is cryptographically detected, but a majority that operates within protocol rules while censoring, re-ordering, or capturing rewards is not misbehaving in a way the protocol can detect. Recovery, when available, depends on social coordination around a user-activated soft fork that repudiates the captured chain, which is exactly the out-of-band political process proof-of-work was designed to avoid.[15]

Proof-of-work reverses the problem. A 51% hashrate holder today faces a market that can respond in several ways at once: new ASICs can be manufactured, offline hardware can be brought online, miners who dislike the attacker's policy can redirect their hashrate to another pool, and entirely new entrants can be drawn in by the profit opportunity an attack creates. The attacker is competing against the

global capacity of the semiconductor industry and the global price of electricity, neither of which is captured by holding the chain. This is what Hugo Nguyen called the timelessness of work: the electricity that secured past blocks cannot be un-spent, and the hardware that secures future blocks is producible by anyone willing to pay.

The empirical record supports the recovery claim even under severe stress. In June 2021, China banned Bitcoin mining, removing roughly two-thirds of the network's hashrate within weeks. Within six months, hashrate had returned to its pre-ban level, redistributed across the United States, Kazakhstan, Russia, and Canada. The attacker in that case was the jurisdiction that had contained most of the world's mining capacity, and the minority recovered by relocating hardware to jurisdictions that had not joined the attack. A proof-of-stake chain in the same position would have had no comparable response: the stake that the Chinese miners held could not have been moved to new validators in other countries, because stake is the chain, not hardware aimed at it.

Economic Incentives for Defense

Miners have invested billions in equipment and infrastructure. This investment is worthless if Bitcoin fails. Miners therefore have strong incentives to defend the network against attacks.

Holders of bitcoin also have incentives to support network health. Running nodes and defending the network against attacks serves holder interests directly.

Axiom of Resistance in the Record

Bitcoin shows the Axiom of Resistance in the record. The network has survived repeated government crackdowns in various countries, bans in some jurisdictions, multiple exchange failures and hacks, sustained negative media coverage, technical attacks on the network, and regulatory uncertainty with hostile legislation.

Since launch, Bitcoin has continued producing blocks approximately every 10 minutes without interruption. The hash rate (computational security) has increased by orders of magnitude.

This empirical track record validates the resistance properties that theory predicts.

Chapter Summary

Bitcoin's sound-money properties are Schelling points upheld by independent validators, not promises made by issuers. Fixed supply, predictable issuance, divisibility, portability, durability, and public verifiability are consensus rules each full node holds locally and validates against every block it receives. Each operator's rules are theirs alone; no one can take them away or force the operator to accept a chain that violates them. A coordinated shift to different rules produces a separate chain that runs alongside the original, and the question of which chain carries monetary value is settled by markets, merchants, and exchanges, not by any threshold of validator participation. Transparent fungibility remains the real base-layer limitation: every unit has a traceable history, and some exchanges reject coins tied to sanctioned or darknet activity. The privacy layers in Chapter 20 exist to restore what the transparent ledger does not give natively. Bitcoin also satisfies the regression theorem through voluntary market adoption, not by decree. Early subjective valuations of censorship-resistant transaction capability provided the non-monetary use from which monetary emergence proceeded, consistent with the praxeological framework of Chapter 9.

Resistance properties and monetary properties are architecturally inseparable. Sound money that can be inflated or frozen by decree is not sound money, and Bitcoin's decentralization, global distribution, and economic incentives for defense ensure that the monetary rules cannot be changed by external pressure. Proof-of-work ties block production to a resource produced outside the ledger, so an attacker who captures a majority of hash power faces a market that can respond by manufacturing more hardware, relocating miners, and drawing in new entrants, the same response that restored hash rate after China's 2021 mining ban. Proof-of-stake systems cannot respond the same way because the security resource

is the coin itself. Resistance is not unlimited. Network-level attacks, mining concentration, and jurisdictional action can disrupt access and degrade security at the margin, and the chapter documents these vectors alongside the resistance properties. What the empirical record supports is that Bitcoin's base layer has continued producing blocks approximately every ten minutes since January 2009 under repeated state and market pressure, which is the operational evidence the theoretical chapters predicted.

Notes

[1]F. A. Hayek, *Denationalisation of Money: The Argument Refined*, 3rd ed. (London: Institute of Economic Affairs, 1990; first published 1976). The epigraph is from Hayek's 1984 lecture "The Future Unit of Value," often quoted as his prediction of something like Bitcoin. Hayek argued that competitive private currencies would discipline governments more effectively than any rule a government could impose on itself, a line of thought that the Bitcoin architecture of Chapter 18 realized twenty-four years later through a different mechanism. For the canonical Austrian sound-money statement this book draws on throughout, see Ludwig von Mises, *The Theory of Money and Credit*, trans. H. E. Batson (New Haven: Yale University Press, 1953; German original 1912), especially part III on monetary policy; cited at Chapter 9, note 4.

[2]Further reading on Bitcoin runs in several directions. For technical depth: Andreas Antonopoulos, *Mastering Bitcoin*, 3rd ed. (O'Reilly, 2023) is the standard developer reference, and Kalle Rosenbaum, *Grokking Bitcoin* (Manning, 2019) is the best illustrated introduction to how the protocol works. For monetary and historical perspective: Saifedean Ammous, *The Bitcoin Standard* (2018), Nik Bhatia, *Layered Money* (2021), and Lyn Alden, *Broken Money* (2023). For narrative history: Nathaniel Popper, *Digital Gold* (2015) on Bitcoin's emergence, and Jonathan Bier, *The Blocksize War* (2021) on the 2015-17 scaling dispute. Primary: Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System* (2008), cited at Chapter 18, note 1; and the Bitcoin Core source tree itself at `https://github.com/bitcoin/bitcoin`. For running infrastructure: Bitcoin Core (`https://bitcoincore.org/`) for full nodes, Sparrow Wallet (`https://sparrowwallet.com/`) for desktop self-custody with coin control, Specter Desktop for multisig, and hardware signers such as Blockstream Jade, Coldcard (`https://coldcard.com/`), and Trezor.

[3]On the Bitcoin Cash fork: the August 2017 hard fork resulted from the block-size scaling dispute that split the Bitcoin development community between 2015 and 2017. Bitcoin ABC (led by Amaury Séchet) and Roger Ver's Bitcoin.com advocacy were the primary drivers of the larger-block chain. Jonathan Bier, *The Blocksize War: The Battle for Control over Bitcoin's Protocol Rules* (self-published,

2021), is the definitive account of the dispute from the perspective of those who opposed the fork. For the Ethereum DAO hack and fork: the DAO (Decentralized Autonomous Organization) was exploited in June 2016 for approximately 3.6 million ETH via a reentrancy attack. The Ethereum Foundation and a majority of validators executed a hard fork in July 2016 to return the funds, producing two chains: Ethereum (ETH, the forked chain) and Ethereum Classic (ETC, the chain that continued the original state). For the technical details of the reentrancy vulnerability, see Phil Daian, "Analysis of the DAO Exploit," *Hacking, Distributed* (June 18, 2016). For the governance implications, see Vitalik Buterin's contemporaneous blog posts at `https://vitalik.eth.limo/` and the Ethereum Foundation announcement of the hard fork at `https://blog.ethereum.org/2016/07/20/hard-fork-completed`.

[4]On the divisibility design, see the early Satoshi Nakamoto–Hal Finney correspondence collected at the Satoshi Nakamoto Institute, `https://satoshi.nakamotoinstitute.org/emails/`, in which Finney raised the question of how finely a bitcoin could be divided and Satoshi described the rationale for the 10^8 (eight-decimal-place) base unit that became the satoshi. The reasoning preserved in those emails has the units sized to give substantial room for both microtransactions and large appreciation in unit value without requiring a future consensus change to the field width.

[5]CoinJoin was proposed by Gregory Maxwell in a BitcoinTalk forum post on August 22, 2013: "CoinJoin: Bitcoin privacy for the real world," `https://bitcointalk.org/index.php?topic=279249.0`. The technique allows multiple users to combine their transactions into a single transaction with multiple inputs and outputs, making it difficult to determine which input corresponds to which output. Implementations include Wasabi Wallet (`https://wasabiwallet.io/`), which uses a coordinator-based equal-output CoinJoin with blind signatures, and JoinMarket (`https://joinmarket.net/`), which uses a maker-taker market for trustless coordination. For the academic treatment of CoinJoin privacy and its limitations under traffic analysis, see Felix Maurer et al., "Anonymous CoinJoin Transactions with Arbitrary Values," *IEEE TrustCom* (2017), and Sjors Provoost and Adam Gibson's work on PayJoin (P2EP), which improves on basic CoinJoin by eliminating the equal-output heuristic. Chapter 20 of this book covers CoinJoin and its variants in detail.

[6]The Lightning Network is a payment-channel network built on Bitcoin enabling high-throughput, low-latency off-chain payments. The protocol was first described in Joseph Poon and Thaddeus Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments" (2016), `https://lightning.network/lightning-network-paper.pdf`. Payments route through a network of bidirectional payment channels secured by Bitcoin's scripting system; only channel open and close transactions settle on-chain. Principal implementations include LND (Lightning Labs, `https://github.com/lightningnetwork/lnd`), Core Lightning (Blockstream, `https://github.com/ElementsProject/lightning`), and Eclair (ACINQ, `https://github.com/ACINQ/eclair`). From a privacy

perspective, Lightning improves fungibility by keeping payment amounts and counterparty identities off-chain, though routing nodes have partial visibility into payment flows; see Bastien Teinturier, "Lightning Network Privacy," https://github.com/t-bast/lightning-docs/blob/master/lightning-privacy.md, for a technical overview. Chapter 20 of this book covers Lightning's privacy model in detail.

[7]Ecash refers to Chaumian blind-signature token systems redeployable over Bitcoin, with Cashu (https://cashu.space/) and Fedimint (https://fedimint.org/) as the primary modern implementations. Cashu uses David Chaum's blind-signature scheme (see Chapter 18, note 2) to issue bearer tokens against a Bitcoin-backed mint; the mint cannot link withdrawal to redemption, restoring the unlinkability that the transparent Bitcoin base layer lacks. Fedimint combines a federated multi-sig custody model with Chaumian ecash issuance, distributing trust across a federation of guardians. For the Cashu protocol specification, see https://github.com/cashubtc/nuts. For Fedimint's design rationale, see Eric Sirion et al., "Fedimint: Federated E-Cash as a Secondary Layer," https://fedimint.org/docs/intro. Chapter 20 covers ecash systems as Bitcoin privacy tools in detail.

[8]For the early BTC-USD exchange that established a continuous market price before the pizza trade, see the BitcoinTalk announcement thread for BitcoinMarket.com, https://bitcointalk.org/index.php?topic=20.0 (March 2010), launched by the user "dwdollar." For the pizza transaction, see Laszlo Hanyecz, "Pizza for bitcoins?", BitcoinTalk forum thread, May 18-22, 2010, https://bitcointalk.org/index.php?topic=137.0, in which Laszlo makes the offer and the user "jercos" (Jeremy Sturdivant) accepts and confirms delivery via Papa John's; the proxy-mediated structure is documented in the thread itself, with Jeremy ordering the pizzas in his own dollars after receiving Laszlo's BTC. For the BitcoinTalk Marketplace board's emergence as the first venue for direct goods-for-bitcoin trade, including the alpaca-sock vendor Michael Stoneman, see the Marketplace subforum archives at https://bitcointalk.org/index.php?board=2.0 (threads from late 2010 onward) and Nathaniel Popper, *Digital Gold* (2015), chs. 4-5, which traces the path from the pizza event through early merchant adoption. For Silk Road's launch and operating model, see Nick Bilton, *American Kingpin: The Epic Hunt for the Criminal Mastermind Behind the Silk Road* (Portfolio, 2017), and the operational descriptions in the federal indictment of Ross Ulbricht (*United States v. Ulbricht*, S.D.N.Y., 2014). The site went live in early 2011 and ran on direct buyer-to-vendor BTC payments routed through a centralized escrow operated by the site.

[9]For regression theorem discussions applied to Bitcoin, see Peter Šurda, "Economics of Bitcoin" (2012); Konrad Graf, "On the Origins of Bitcoin" (2013); Eric Voskuil, "Regression Fallacy," *Cryptoeconomics* (2020). These sources are collected at Chapter 9, note 5, which covers the regression debate in full.

[10]On BGP hijacking and network-level attacks against Bitcoin, see Maria Apostolaki et al., "Hijacking Bitcoin: Routing Attacks on Cryptocurrencies," *IEEE Symposium on Security and Privacy* (2017). For eclipse attacks, see Ethan Heilman

et al., "Eclipse Attacks on Bitcoin's Peer-to-Peer Network," *USENIX Security* (2015). For the stealthier EREBUS attack, see Muoi Tran et al., "A Stealthier Partitioning Attack against Bitcoin Peer-to-Peer Network," *IEEE S&P* (2020).

[11]China's ban on Bitcoin mining was announced by the National Development and Reform Commission (NDRC) and the People's Bank of China (PBoC) in May 2021, with enforcement intensifying through June 2021. The ban caused Bitcoin's network hashrate to drop by approximately 50% over the following weeks. By early 2022, hashrate had recovered to and surpassed pre-ban levels, with mining activity redistributed primarily to the United States, Kazakhstan, and Russia. For the regulatory text, see NDRC, "Notice on Rectifying Virtual Currency 'Mining' Activities" (May 2021). For the hashrate recovery data, see Cambridge Centre for Alternative Finance, "Cambridge Bitcoin Electricity Consumption Index," mining map, `https://ccaf.io/cbnsi/cbeci/mining_map`, and Nic Carter, "How Much of Bitcoin Mining Is Powered by Clean Energy?" *CoinDesk* (2021), which documents the geographic redistribution. The episode is the primary empirical test of Bitcoin's resistance to state-level mining suppression cited throughout this chapter.

[12]The Tor Project (`https://www.torproject.org/`) develops and maintains the Tor anonymity network, which routes TCP traffic through a series of volunteer-operated relays using onion encryption so that no single relay knows both the origin and destination of a connection. Bitcoin Core has supported Tor natively since version 0.12 (2016), and nodes run as hidden services (`.onion` addresses) are not reachable from the clearnet, hiding their IP addresses from peers and observers. For the Tor integration in Bitcoin Core, see the documentation at `https://github.com/bitcoin/bitcoin/blob/master/doc/tor.md`. For the privacy and network-topology implications of Tor usage for Bitcoin nodes, see Ethan Heilman et al., "Eclipse Attacks on Bitcoin's Peer-to-Peer Network," *USENIX Security* (2015), already cited at note 6, which discusses both the vulnerability and Tor-based mitigations.

[13]Mining pool concentration data: Cambridge Centre for Alternative Finance, "Cambridge Bitcoin Electricity Consumption Index," pool distribution data, `https://ccaf.io/cbnsi/cbeci/mining_map`; real-time pool hashrate distribution tracked at `https://mempool.space/mining`.

[14]On Stratum V2, see the protocol specification and reference implementation maintained by the Stratum V2 Working Group, `https://stratumprotocol.org/`, particularly the Job Declaration role that lets a miner or mining farm construct its own block templates while continuing to participate in pool payout via a separate share-tracking channel. For Datum, see the project documentation from Ocean Mining, `https://ocean.xyz/docs/datum`, which implements decentralized template construction for pool members so that the pool operator no longer chooses which transactions appear in candidate blocks. Both protocols leave the share-aggregation and payout layer with the pool while moving template construction to the individual hashing endpoint, which is the architectural

separation that addresses pool-operator censorship without requiring miners to leave the pool's variance-smoothing payout structure.

[15]For the resource-externality argument developed here, see Hugo Nguyen, "Proof-of-Stake & the Wrong Engineering Mindset" (2018), and "Work is Timeless, Stake is Not" (2018). For the canonical proof-of-stake critique, see Andrew Poelstra, "On Stake and Consensus" (2015). Vitalik Buterin's counter-arguments are developed in "Proof of Stake: How I Learned to Love Weak Subjectivity" (2014) and "Why Proof of Stake" (2020); Buterin concedes that recovery from majority capture in proof-of-stake requires coordinated social forking. For the China-ban recovery data, see the Cambridge Centre for Alternative Finance, "Cambridge Bitcoin Electricity Consumption Index," mining map, `https://ccaf.io/cbnsi/cbeci/mining_map`.

20

BITCOIN PRIVACY AND MONETARY LAYERS

> "Bitcoin itself cannot keep a user anonymous; external IP tracking and transaction analysis can reveal identities."
>
> Hal Finney, Bitcoin Forum (2010)

Chapter 19 established Bitcoin as sound and resistant money.[1] Those strengths do not make Bitcoin private by default. This chapter turns to Bitcoin's actual privacy model. It begins with the attacks that make privacy necessary, then turns to the primitives and the layers built to defeat them.

The point is neither maximalist evangelism nor dismissal. Bitcoin's base layer exposes real information, and the chain-analysis industry has spent more than a decade building tools to exploit that exposure. Privacy on and around Bitcoin comes from understanding what those tools can see, then applying layered design and disciplined tool choice against them. The chapter proceeds in that order: first the attacks against base-layer privacy, then the script primitives that higher layers depend on, then CoinJoin, PayJoin, and CoinSwap at the base layer, Lightning, Spark, and Ark off-chain, Chaumian ecash with its custodial tradeoff, and finally the client-side validation designs that push privacy further at the cost of maturity.

20.1 How Base-Layer Privacy Fails

A Public Ledger by Design

Bitcoin's blockchain is public. Every transaction ever made is visible to anyone. Amounts, addresses, and timing are all recorded permanently, and the permanence is the point: it is what allows nodes to verify the ledger without trusting any operator. The same property that enables trustless verification also underwrites the privacy attacks that follow. Every heuristic in the chain-analysis toolkit is a way of squeezing information out of data that the protocol publishes by necessity.

The attacks compose. Individually, most are probabilistic; together, they collapse the anonymity set toward a point. Understanding them in roughly the order of severity is also the order of the historical arc: address reuse was the first documented leak, the common-input-ownership heuristic became the industrial workhorse, change detection extended clusters forward in time, timing and network analysis added an IP layer, and KYC linkage anchored the whole graph to real identity.

Address Reuse: The Foundational Leak

An address reused across multiple payments collapses every associated output into a single cluster without any further analysis. All UTXOs held at that address are controlled by the same private key by definition. Every payment into and out of the address is publicly joined on the blockchain forever. Reuse also chains outward: whoever paid the address, whoever the address paid, and every observer of the chain obtains a full ledger of the owner's activity at that location.

Modern hierarchical deterministic wallets, specified in BIP32 in 2012 and extended through BIP44, BIP84, and BIP86, eliminate the most common pathway to reuse by deriving a fresh address from a single seed on every receive and routing change to a separate derivation path. The UX problem the specifications solved was backup durability: pre-hierarchical deterministic wallets kept a finite reserve of pregenerated keys, so an old backup became stale

after the wallet exhausted that reserve. Privacy gains followed almost as a side effect.[2] Reuse persists, however, in categories that sit outside normal wallet flows: donation addresses embedded in forum signatures, GitHub READMEs, and social profiles; mining pools that pay miners to fixed addresses; older exchange deposit addresses; paper wallets and single-address brain wallets; and ransomware payment addresses that attackers deliberately reuse across victims. Each reused address is a cluster anchor that the rest of the chain-analysis toolkit can grow outward from.

The Common-Input-Ownership Heuristic

If a transaction has more than one input, the inputs are presumed to be controlled by the same entity. The reason is cryptographic: spending a UTXO requires a valid signature for its scriptPubKey, so combining multiple inputs in one transaction requires simultaneous access to all the corresponding private keys. Absent deliberate coordination between parties, the only reason to combine inputs is that one wallet holds all of them.

The heuristic appears in the Bitcoin whitepaper's own discussion of privacy and was formalized against the full chain in Meiklejohn et al., *A Fistful of Bitcoins*, in 2013.[3] The paper used the common-input-ownership heuristic together with change detection to collapse roughly twelve million addresses into about 3.4 million clusters. It remains the workhorse of the blockchain-analysis industry, applied now at chain scale by every major compliance vendor. The heuristic fails when a transaction's inputs come from different parties, which is the entire design goal of CoinJoin and PayJoin. In practice, however, most transactions are not CoinJoins, most users never use privacy tools, and a single KYC touchpoint anywhere in a cluster's history attaches a real-world label to the whole cluster forever.

Change Detection

A one-input-two-output transaction, which is the ordinary consumer pattern, has one output paying the counterparty and one

output returning change to the sender. Identifying which is which lets the analyst extend the sender's cluster forward in time by treating the change output as the seed of the next transaction. Stacked with the common-input-ownership heuristic, correct change detection produces the "peel chain" pattern that drives transitive cluster growth across the chain.[4]

Several heuristics stack to make this work. The strongest is the reused-address heuristic: if one output goes to an address that has been used before and the other goes to a fresh address, the reused one is almost certainly the payment and the fresh one is change, because wallets generate change addresses automatically while payment addresses are defined by the recipient manually. Wallet fingerprinting identifies which output was generated by the same software that built the transaction by matching script type, ordering conventions, and behavioral quirks.

The round-number heuristic notes that humans pay in round amounts, so an output equal to one Bitcoin or to a hundred dollars at the current exchange rate is almost always the payment, and the non-round leftover is change. The unnecessary-input heuristic notes that rational coin selection never includes inputs that are not needed to cover the payment; if one output could have been funded with a strict subset of the actual inputs, that output is change and the other is the payment. Script-type matching observes that wallets usually spend and change in the same script type, so a transaction whose inputs are all P2WPKH and whose outputs are one P2WPKH and one older type marks the P2WPKH output as change. Replace-by-fee behavior marks change too: when a wallet bumps fees on an unconfirmed transaction, it almost always pays the extra fee by reducing the change output amount, so the output whose value shrinks between two mempool versions is change.

No single one of these is decisive. In combination, they identify change correctly often enough to keep the peel chain extending. The analyst needs to be right only most of the time.

Timing and Network Analysis

Chain-level heuristics operate on what the ledger records. Timing and network analysis operate on what happens around the ledger, before and after confirmation.

The most direct network-layer attack runs many Bitcoin full nodes as listening posts on the peer-to-peer network. When a new unconfirmed transaction appears, the listening nodes record which peer sent them the first notification. Under honest broadcast behavior, the first peer to announce a transaction is likely the originating node, which ties the transaction to an IP address. Pre-2015 Bitcoin Core broadcast new transactions to all peers simultaneously, which made this attack close to trivial. Modern Bitcoin Core uses randomized per-peer exponential delays that weaken but do not defeat the attack when the adversary holds many vantage points. The LinkingLion entity, documented by the researcher 0xB10C in 2023, has operated this kind of attack against Bitcoin and Monero from specific autonomous systems since at least 2018, opening short-lived listening connections to reachable nodes with a rotating pool of fake user agents.[5] Dandelion++, a defense that propagates transactions through a random stem of relays before broadcasting normally, has been deployed in Monero since 2020 but has not been merged into Bitcoin Core.

Timing also leaks within the ledger itself. Two transactions broadcast within milliseconds of each other probably share an originator, diurnal activity patterns leak time zone and therefore geographic region, and sustained automation with low jitter identifies mining-pool payouts, exchange rebalancers, and arbitrage bots. Cross-chain bridge deposits and withdrawals correlate by amount and characteristic delay. Lightning payments leak routing information through HTLC hold times and through probing attacks that learn channel balances by sending deliberately invalid payments. None of these is decisive on its own. In combination with on-chain heuristics, they collapse the anonymity set further.

KYC Anchors and the Analysis Industry

The attacks above produce clusters of addresses. Turning a cluster into a person requires an anchor to a real identity, and the chain-analysis industry has built its business around collecting those anchors. Every major centralized exchange collects know-your-customer information under anti-money-laundering regulations: legal name, address, government ID, sometimes phone and banking details. Deposit and withdrawal addresses tied to KYC'd accounts are labeled inside the exchange's internal systems and shared with compliance vendors through commercial contracts, with governments through subpoenas and mutual legal assistance treaties, and with outside researchers through deposit-address harvesting. One KYC touchpoint anywhere in a cluster's history labels the cluster's real-world owner for good. The FATF Travel Rule, adopted in 2019 and rolled out globally through 2024, requires virtual-asset service providers to share originator and beneficiary information for transfers above certain thresholds, extending the KYC linkage across the industry.[6]

Chainalysis, Elliptic, TRM Labs, Mastercard's CipherTrace, and a long tail of smaller vendors sell the resulting capabilities to banks, exchanges, governments, and law-enforcement agencies.[7] The tooling ranges from real-time transaction scoring to forensic investigation platforms. The probabilistic nature of the underlying heuristics is often understated in marketing and in court; the Bitcoin Fog prosecution of Roman Sterlingov in 2024 became a flashpoint over whether chain-analysis attribution methodology meets ordinary standards for scientific evidence.[8] The industry has also opposed Bitcoin privacy upgrades when those upgrades threatened its product. Its incentives are what they appear to be.

Wallet Sync: Where the Wallet Meets the Chain

The attacks above describe what an adversary sees when they read the chain. Before the adversary reads the chain, the user's wallet has to read it first, and every query a wallet makes is a message to some node about which addresses matter to the user.

The wallet's sync architecture is the chokepoint where on-chain privacy meets network-layer privacy, and different architectures leak different amounts to whoever the wallet is talking to.

A **self-hosted full node** is the privacy-maximal architecture. A validating node downloads every block since genesis (∼700 GB of storage, ∼80–100 GB growth per year, 6–24 hours of initial sync on consumer hardware) and keeps the UTXO set locally. The wallet queries the node over localhost RPC; no remote party learns which addresses the wallet watches, which outputs it holds, or which transactions it is waiting for. The only outbound signal is transaction broadcast, a single message that can be wrapped in Tor. Pruned nodes (reduces storage to ∼5–10 GB) preserve full validation while dropping historical blocks from disk, so the privacy properties hold without the full archival footprint. The standard sovereign setup runs Bitcoin Core as a background service with a desktop wallet (Wasabi, Sparrow, or Core's own wallet) talking to it locally; all history sync, fee estimation, and balance queries go through the local node.[9]

Electrum and Electrum-protocol wallets use a different architecture: the client sends address-level queries to an Electrum server which indexes the entire chain by script-hash. The server returns UTXOs and history directly in seconds. The privacy posture is the worst of the practical architectures because every address the wallet watches arrives at the server in a single connection; the server learns the full wallet cluster, balance, history, and IP address at once. The remedy is to run the Electrum server yourself on the same host as your full node. ElectrumX (Python), electrs (Rust), and Fulcrum (C++) are the three standard self-hosted implementations, with electrs shipping as a default component in several sovereign-node distributions. Self-hosted Electrum preserves the fast wallet UX while removing the third-party leak: when the server is the user, no one else sees anything.

BIP 37 bloom filters, introduced by Mike Hearn and Matt Corallo in 2012, were the first light-client protocol on Bitcoin's peer-to-peer network. The design was simple: the wallet constructs a Bloom filter over its addresses, sends it to a serving full node, and

receives only the transactions that match. The privacy claim was that the filter's false-positive rate would obscure which addresses the wallet cared about. The privacy claim was false. Gervais, Karame, Gruber, and Capkun showed in 2014 that a serving node observing even one filter could substantially narrow the wallet's address set, and that intersecting filters from the same wallet recovered addresses with near-certainty. Jonas Nick's 2015 measurement against live BitcoinJ deployments confirmed that the effective attack recovered nearly all of a wallet's pubkeys from one observed filter. Bitcoin Core disabled BIP 37 serving by default in release 0.19.0 (November 2019); the protocol was never formally deprecated, but it is no longer served by the majority of the network.

BIP 157/158 compact block filters, specified by Olaoluwa Osuntokun, Alex Akselrod, and Jim Posen in 2017, inverted the direction of the leak. Instead of the wallet sending a filter to the server, the server computes a deterministic Golomb-coded filter (roughly 20–30 KB per block) covering every scriptPubKey in the block, and every wallet downloads the same filter and matches locally. The server never sees what the wallet is watching; the filter is the same for everyone. The residual leaks are network-level (the wallet's IP remains visible) and block-request-timing (on a filter match the wallet downloads the full block, and requesting rare blocks from a single peer reveals which blocks contain transactions of interest). Downloading matched blocks from rotating outbound peers mitigates the second attack; Tor handles the first. Bitcoin Development Kit (BDK) exposes the filter primitives to wallet developers, and the protocol is served by `btcd` and by Bitcoin Core's compact block filter index.[10]

Wasabi Wallet, launched by zkSNACKs in August 2018, pioneered the use of client-side block filtering in a consumer Bitcoin privacy wallet. Before v2.6.0, Wasabi served custom filters from its own backend because BIP 157/158 was not yet standardized or widely deployed when Wasabi's architecture was built. Those filters used a specific construction including only native SegWit script types, which made them smaller in practice than full BIP 158 block filters covering every scriptPubKey. Wasabi migrated to

standard BIP 158 filters in v2.6.0 (May 2025), which also lets users point the wallet directly at their own Bitcoin Core node without any intermediate server.

The resulting tier structure tracks how much the wallet leaks to its serving peers. A full node leaks nothing address-specific. A BIP 157/158 client leaks block-match timing and IP. A self-hosted Electrum server leaks nothing to the outside world while keeping Electrum's UX. A third-party Electrum server leaks the entire wallet cluster. A BIP 37 client, if one still existed in production, leaks nearly as much as third-party Electrum. The order reflects what the wallet's sync channel tells its counterparty about the user, and it is the order in which wallet architectures should be preferred.

Miner-Layer Protocol Changes

Stratum V2, the successor to the mining pool protocol Bitcoin has used since 2012, is primarily a security upgrade to the miner-to-pool channel. The V1 protocol runs cleartext over TCP, which leaks the work stream to any network observer between the miner and the pool and lets an in-path adversary redirect hashrate through hashrate-hijacking attacks that were documented repeatedly across the 2010s. V2 encrypts the entire miner-to-pool connection using the Noise framework with an AEAD cipher, which closes the in-path attack class and the passive-observation leak together. On top of that security base, V2's optional Job Declaration extension lets the miner construct their own block templates; the pool no longer chooses which transactions enter pooled work, so a pool operator who tried to censor specific transactions in pooled hashrate would lose the hashrate to a self-declaring miner.

Braiins Pool, DEMAND Pool, and Public Pool are the production Stratum V2 deployments at the time of writing. OCEAN Pool implements the parallel DATUM protocol, which uses a different mechanism (a transparent payment-settlement record on the pool's own Bitcoin transactions) to achieve the same non-custodial and censorship-resistant properties V2's Job Declaration targets. Adoption of either mechanism is incomplete, and the adoption curve

matters for the censorship properties of the chain more than for any single miner's privacy.[11]

Dandelion++, the transaction-propagation defense Monero deployed in 2020, has not been merged into Bitcoin Core because the design introduces denial-of-service vectors at the stem relays that the Bitcoin Core review process has not found acceptable. The practical alternative adopted by privacy-focused Bitcoin wallets is to broadcast each transaction under a fresh Tor identity: the wallet opens a new Tor circuit, connects to a single randomly chosen Bitcoin node it has never connected to before, sends the transaction, and tears down both the connection and the circuit. Wasabi Wallet introduced this pattern and Bitcoin Core now supports it through its `-onlynet=onion` and per-connection circuit isolation options. The remaining attack surface is timing correlation between the wallet's Tor circuit and the first peer's gossip to the rest of the network, which weaker adversaries cannot exploit but a global passive observer still can. The structural point across these protocols is that network-layer privacy on Bitcoin is an engineering program that has produced partial gains the wallet's operator must opt into.

What the Base Layer Therefore Does Not Provide

Taken together, the attacks explain why the base layer does not provide transaction confidentiality (amounts are visible), sender-receiver unlinkability (addresses are visible and clusterable), denial that a transaction occurred (every transaction is public), or protection from chain analysis (the patterns are analyzable by anyone with the resources). The layers built on top of the base layer exist to defeat specific combinations of these attacks, and each layer does so at a specific cost. The rest of the chapter examines what each one accomplishes and what it leaves unchanged.

20.2 Programmable Money: Script Primitives

Bitcoin Script

Each transaction output contains a script specifying the conditions that must be satisfied to spend it. Bitcoin Script is deliberately non-Turing-complete and cannot loop or access external state, which makes scripts predictable and auditable and limits programmability to financial primitives.

Multisignature

Multisignature (multisig) scripts require multiple keys to authorize spending. A 2-of-3 multisig requires any two of three designated keys, enabling shared control without trusting any single party. Legacy multisig pays for that control with on-chain disclosure: the script reveals every participant public key, and the spend reveals which keys signed. Schnorr signatures make the cleaner version practical. MuSig2 and FROST move the threshold logic into an aggregate signing protocol, so the chain sees one public key and one signature while the participant set and threshold remain off-chain.[12] Threshold ECDSA exists too, but the protocols are more complex and lack the same clean security proof. Corporate treasuries, inheritance schemes, exchange hot wallets, and escrow arrangements all use the primitive to replace a trusted custodian with programmable shared custody.

Timelocks

Timelocks prevent spending until specified conditions are met. Two types exist:

Absolute timelocks (CLTV, CheckLockTimeVerify) prevent spending until a specific block height or timestamp. A transaction locked until block 900,000 cannot be spent before that block, regardless of who holds the keys.

Relative timelocks (CSV, CheckSequenceVerify) prevent spending until a specified time has elapsed since the output was created.

An output locked for 144 blocks cannot be spent until 144 blocks (approximately one day) after its creation.

Timelocks enable time-delayed transactions, vesting schedules, and, critically, the revocation mechanisms that make payment channels secure.

Hash Locks

Hash locks require revealing a secret value (preimage) whose hash matches a specified hash. The script contains a hash; spending requires providing the preimage.

This seems simple but enables powerful constructions. If Alice wants to pay Carol but has no direct channel, she can route through Bob using hash locks: Alice pays Bob conditional on Bob revealing a secret that only Carol knows. Bob pays Carol the same way. When Carol reveals the secret to claim her payment from Bob, Bob learns the secret and can claim his payment from Alice. The payment atomically succeeds or fails across the entire path.

Hash Time-Locked Contracts (HTLCs)

Combining hash locks with timelocks creates Hash Time-Locked Contracts (HTLCs). An HTLC can be spent in two ways: by revealing the hash preimage (success path), or after a timeout expires (refund path).

HTLCs are the foundation of the Lightning Network. They enable trustless multi-hop payments: either the payment completes across all hops when the final recipient reveals the preimage, or all participants recover their funds after timeout. No intermediate routing node can steal funds or block payments indefinitely.

What These Primitives Do

Together these primitives transform Bitcoin from simple digital gold into programmable money. Multisig replaces trusted custodians with shared-key control; timelocks replace trusted schedulers with block-height enforcement. Hash locks and HTLCs go further:

they enable conditional, multi-hop payments that no escrow agent needs to mediate.

Taproot (BIP 341, activated November 2021) adds a privacy benefit to the script-primitive stack that older output types could not offer.[13] Every Taproot output commits to a Merkle tree of alternative spending scripts, and a spender who uses any one leaf reveals only that leaf. The other branches of the tree remain unpublished and indistinguishable from bytes that were never written. A 2-of-3 multisig with a 1-of-1 cooperative path and a timelocked recovery path spends on-chain as a single Schnorr signature when the cooperative path is taken, with the other branches never appearing. Chain analysis cannot tell from the spending transaction whether the output was a simple single-sig, a complex multi-party script, or anything in between. The script-path visibility rule turns the expressive power of Bitcoin script from a privacy cost into a privacy asset: complexity that exists only to be used occasionally is hidden by default, and only the path exercised is disclosed.

Each primitive removes a category of trusted intermediary. Together they enable Layer 2 systems like Lightning that inherit Bitcoin's settlement security while gaining speed and stronger privacy.

20.3 Privacy Solutions: Collaborative Transactions and Swaps

CoinJoin: Collaborative Transactions

CoinJoin combines inputs from multiple users into a single transaction with multiple outputs.[14] Participants are not paying each other; each person sends funds to themselves. The privacy gain comes from breaking the link between inputs and outputs: if Alice, Bob, and Carol each contribute one input and each receive one output of equal value, an observer cannot determine which output belongs to which input.

CoinJoin requires coordination between participants. Two main approaches have emerged.

JoinMarket, the earlier implementation, uses market-based coordination: "takers" initiate CoinJoin rounds and pay "makers" for

liquidity.[15] Each taker coordinates its own transaction, distributing coordination across many participants instead of centralizing it in a single service. The transaction has one equal-denomination output for each participant, which is the output set that receives the anonymity. A taker can also make a payment inside the CoinJoin by adding a separate payment output while preserving the equal-denomination mix output; makers provide liquidity and earn fees, but they cannot direct a payment output because they do not control the transaction template. Takers must run more complex software and pay for participation, but no external coordinator can be shut down.

Wasabi 1.0 used a similar equal-denomination model with a central coordinator and Chaumian blind signatures to enable anonymous registration of the equal amount output. Wasabi 2.0 changed the construction with WabiSabi, which uses keyed-verification anonymous credentials (KVACs) to issue blind signatures over arbitrary amounts. Participants can register arbitrary-value inputs and choose output values that fit their wallets instead of forcing every private output into one denomination. By avoiding repetitive equal chunks, WabiSabi spends less block space and improves both wallet efficiency and privacy.

It also generalizes CoinJoin payments: any participant can register payment outputs during a round, and the coordinator cannot distinguish a payment output from a self-transfer output. This aggregation produces much larger transactions with higher anonymity sets than JoinMarket's distributed model typically achieves. The tradeoff is that static coordinators present censorship challenges: a coordinator can be taken offline through legal pressure or technical attack. The WabiSabi protocol is open; anyone can run a coordinator, and multiple coordinators can operate simultaneously. When Wasabi's original coordinator ceased operation due to regulatory pressure, alternative coordinators showed that the protocol itself was not tied to one operator, even if liquidity and usability still depended on implementation details.[16]

PayJoin: Payment-Integrated Coinjoining

PayJoin is designed explicitly for payment transactions. Instead of Alice sending Bob a payment in a simple transaction, both Alice and Bob contribute inputs. The result looks like an ordinary payment, not a CoinJoin.

PayJoin operates on a different privacy model than CoinJoin. Instead of creating anonymity through mixing with other users, PayJoin is steganographic: it makes privacy-enhancing transactions indistinguishable from ordinary transactions. CoinJoin's equal-amount outputs create an obvious fingerprint on the blockchain; PayJoin transactions look like any normal payment.

The privacy benefit is structural. PayJoin breaks the common input ownership heuristic, the assumption that all inputs in a transaction belong to a single entity. Chain analysis depends heavily on this heuristic. When PayJoin transactions are indistinguishable from regular transactions, analysts cannot know which transactions violate the heuristic, degrading the reliability of clustering analysis across all transactions, not just PayJoin ones. Wider PayJoin adoption thus improves privacy for the entire network by introducing uncertainty into chain analysis assumptions.

PayJoin also hides the payment amount from the chain's plain arithmetic. In a normal payment, the recipient output is the payment amount. In a PayJoin, the receiver adds their own input value on top, so the output they receive is no longer equal to the sender's payment. The protection is conditional: if an adversary later separates sender-owned inputs from receiver-owned inputs, the actual payment amount can be reconstructed. The same construction can improve block-space efficiency. A receiver can add unrelated payments to other people into the same transaction, or coordinate additional participants' inputs and outputs, turning the receiver into the coordinator of a multiperson CoinJoin around the payment. That saves block space by combining several economic transfers into one transaction, but it does not have the coordinator privacy like WabiSabi; the receiver coordinating the transaction can see the participant roles they arrange.

Limitations

CoinJoin and PayJoin improve privacy but have distinct limitations. CoinJoin requires other participants, and the effective anonymity depends on the liquidity available in the round. WabiSabi removes the fixed-denomination constraint that made older coordinator designs produce obvious change, so unequal amounts are not the core limitation. The remaining constraints are operational: a coordinator can deny service, set admission policy, choose round parameters, or disappear under legal pressure; one coordinator tends to attract most liquidity, making that coordinator the practical meeting point even when the protocol permits many coordinators. Rounds also depend on participant liveness. A participant who fails to sign can force a blame round, and Sybil liquidity can reduce effective anonymity if an adversary pays the fees to join with many coins. After the CoinJoin, careless spending can still undo the privacy by merging histories that the round separated.

PayJoin's limitations are different in kind. The steganographic claim that PayJoins look like ordinary payments is conditional. The unnecessary-input heuristic, analyzed against PayJoin by Adam Gibson and LaurentMT in 2018 and formalized by Ghesmati, Fdhila, and Weippl in 2022, flags transactions whose inputs exceed what rational coin selection would have produced; LaurentMT's measurement found roughly a third of two-output transactions violating one of the variants, so PayJoins that fall into that third are statistically detectable.[17] Wallet fingerprints break the claim along separate axes: mixed input script types in particular announce multi-party origin so reliably that BIP 78 disallowed them entirely until September 2024, and sequence values, nLockTime, BIP 69 input ordering, and signature R-value patterns survive the join with their original fingerprints intact.

The construction also exposes one of the receiver's UTXOs to the sender by design, which a malicious sender can exploit as a deanonymization channel against the receiver they are paying; the probing variant against a public merchant endpoint is bounded by the fact that the sender's first message in BIP 78 is a fully signed and broadcastable original transaction, so a sender that aborts to

probe risks the receiver broadcasting that fallback and pocketing the payment. Operationally, the receiver pays the fee for the input they contribute, replace-by-fee bumping requires both parties to coordinate again because both signed the transaction, and BIP 78's default output-substitution setting lets a compromised PayJoin server redirect payment unless the sender opts out.

Infrastructure cost has been the binding constraint on adoption: BIP 78 required the receiver to host a TLS-protected HTTP endpoint reachable at payment time, and BIP 77's async successor (merged in May 2025) lifts that requirement only by routing traffic through an OHTTP directory whose sole production deployment, payjo.in, becomes a new trust principal.[18] Six years after BIP 78, production sender-and-receiver implementations number in the single digits, Bitcoin Core has no native integration, and the chain-analysis-uncertainty benefit the steganographic argument relies on is therefore real but small.

Both tools also require users to understand and correctly apply them, adding implementation complexity on top of the protocol-level limitations on each side.

CoinSwap: Cross-Transaction Unlinkability

CoinSwap is the other on-chain approach to severing the transaction graph. Where CoinJoin combines multiple users' inputs and outputs into a single transaction, a CoinSwap has two parties exchange outputs across separate transactions: Alice's coin goes to Bob's address in one transaction, and Bob's coin goes to Alice's address in another. An outside observer sees two ordinary-looking sends. No transaction fingerprint distinguishes a CoinSwap from a pair of unrelated payments, because each transaction has one input and two outputs and carries no on-chain coordination signal. The link between the coin that was spent and the coin that was received disappears across multiple hops: chain analysis can follow a coin forward to a swap counterparty's address but cannot determine whether the value continued in that direction or whether the counterparty's subsequent spend represents the swapped coin or an unrelated one.

The privacy property is plausible deniability at scale. With CoinJoin, chain-analysis tools can identify the transaction as a mixing event, even if they cannot determine which output belongs to which participant. With CoinSwap, they face a more difficult problem: there is no mixing event to identify. The swap looks like any other spend, and a chain of swaps through multiple counterparties produces a transaction graph where the economic flow is undetectable, not ambiguous in the way a CoinJoin is ambiguous.

The construction is mainly theoretical on Bitcoin at present. The atomic swap protocol, in which a hash-timelock contract enforces that both legs of the exchange either complete or both revert, dates to Tier Nolan's 2013 proposal on the Bitcoin Talk forum.[19] Same-chain CoinSwaps for on-chain privacy have been explored in research and in early prototype tooling, but no widely deployed implementation exists with meaningful usage. The demand-side coordination problem is the binding constraint: finding a willing counterparty with the right output size at the right time requires either a public market (which leaks information about intent) or a trusted meeting point (which reintroduces a coordination server).

Lightning is the off-chain instantiation of the same primitive. Each Lightning payment hop is an atomic CoinSwap between adjacent channel partners, enforced by Hash Time Locked Contracts on the same elliptic-curve assumptions Bitcoin uses for spending. For an Alice-to-Carol payment through Bob, Alice and Bob execute one swap while Bob and Carol execute another; neither Bob nor Carol can link Alice's incoming HTLC to Carol's outgoing payment without additional information. The construction pushes CoinSwap coordination off the base layer entirely, where it executes at millisecond latency with no on-chain footprint per payment. Lightning is, in this sense, the production deployment of the CoinSwap model, shifted into the channel-graph layer.

20.4 Non-Custodial Off-Chain Payments

Lightning, Ark, and Spark are the three off-chain constructions examined here that move Bitcoin payments off-chain while preserv-

ing each user's ability to unilaterally return to the base layer without counterparty cooperation. They differ in how liquidity is managed, what the operator sees, and what scripting capability the construction exposes, but all three leave custody with the user. Custodial constructions are the subject of the next section.

Lightning Network: Payment Channels

The Lightning Network enables off-chain transactions through payment channels.[20] Two parties lock bitcoin in a multisignature address, then exchange signed transactions updating the balance between them without broadcasting each update to the blockchain. As long as the two parties continue to cooperate, the channel can handle arbitrarily many payments; only the eventual close settles back to the base layer. Channels can also be linked, so that Alice pays Carol through Bob whenever channels exist between Alice-Bob and Bob-Carol, and this routing extends the payment graph far beyond the set of direct channel partners. A small number of well-connected nodes can therefore support a large number of payments among parties that never open channels directly with each other.

Instant Finality

Base-layer confirmation is probabilistic and slow: blocks arrive on average every ten minutes, intervals vary widely, and prudent recipients wait for additional confirmations against reorg. Lightning replaces that wait with an immediately enforceable claim inside the channel structure, a signed commitment transaction the recipient can broadcast if needed. Settlement is cryptographic within the channel instead of probabilistic on the chain: either the recipient reveals the preimage and the payment succeeds across all hops atomically, or timeouts expire and funds return to senders. Point-of-sale, micropayments, and machine-to-machine flows become practical once payment overhead collapses to milliseconds.

Privacy Properties and the Channel-Graph Model

Lightning offers privacy properties the base layer lacks. Payments within channels are invisible on the blockchain; only channel opening and closing appear on-chain. Onion routing means each intermediate node learns only its predecessor and successor, not the full path. Announced channel capacities are visible, but the balance between the two channel parties is private.

The privacy model differs from base-layer Bitcoin instead of improving on it in a single direction. The public channel graph reveals network structure while payment paths through it remain private. Routing nodes see the amount passing through them, but payments can be split across multiple paths so no single routing node necessarily sees the full amount. Topology stays public while actual usage stays hidden.

Privacy Limitations

Lightning's privacy properties are weaker than they might appear.[21]

Balance discovery through probing is a serious vulnerability. An adversary can send probe payments through target channels, observing which amounts succeed and fail. Research has shown efficient balance estimation under specific network conditions, often at low cost to the attacker because probes are designed to fail. Through systematic probing, adversaries can infer the balance distribution in channels they can route through.

Routing nodes observe payment amounts passing through them. While onion routing hides the payment's origin and destination, intermediate nodes see values. A routing node that appears in many paths gains statistical information about network payment flows.

Lightning Service Providers (LSPs) present particular privacy concerns. Many mobile wallets connect to a single LSP for channel management and routing. The LSP learns the user's node identity and network address, along with much of the payment activity if payments route through the LSP's node. This reintroduces the trusted third party that Bitcoin was designed to eliminate.

On-chain channel opening and closing transactions link Lightning activity to base layer transactions. Clustering analysis can connect these to other user transactions. Force-close transactions can also reveal channel states at closure time.

One 2021 cross-layer deanonymization study found that combining on-chain and Lightning data could link 43.7% of Lightning nodes in its dataset to associated Bitcoin addresses.[22] The interaction between layers creates information leakage that neither layer alone would reveal.

Lightning Limitations and Current State

Lightning has operational limitations beyond privacy. Channels require locked capital, receiving capacity requires counterparties to lock funds, payments require online presence or watchtower services, and optimal operation requires liquidity management across channels. Large payments may fail to find paths with sufficient liquidity.

Privacy improvements are actively being developed and deployed. Route blinding, especially in newer BOLT 12-style flows, can improve receiver privacy by hiding node identity and parts of the path from counterparties, though support remains implementation-dependent. Payment splitting across paths reduces information available to any single routing node. Trampoline routing can further reduce how much topology some senders or routing intermediaries learn, but its privacy effects depend on how it is implemented. The Lightning protocol undergoes continuous development with regular specification updates; many privacy concerns identified by researchers are being addressed through protocol improvements, though some, such as balance probing, remain structural challenges inherent to the routing discovery process. Lightning remains developing technology whose privacy properties shift with implementation choices and network conditions alike.

Ark: Shared-UTXO Payment Rounds

Ark, specified by Burak Keçeli starting in 2023, is an off-chain construction that removes Lightning's liquidity-management burden at the cost of a different trust model.[23] Instead of opening a channel with a specific counterparty and managing inbound and outbound capacity, a user joins an Ark round by handing bitcoin to a shared UTXO that the Ark server controls, receives virtual UTXOs inside that shared UTXO whose value is immediately spendable within the Ark protocol, and settles to a new shared UTXO at the end of each round. Users can receive payments without opening channels in advance, which is the usability gap Lightning has not closed for mobile wallets whose receiving capacity is priced and provisioned through Lightning Service Providers.

The trust model is different from Lightning's. A misbehaving Lightning counterparty cannot steal the channel's funds because the revocation-key construction lets the honest side unilaterally close on any stale channel state they publish. An Ark server likewise cannot steal user funds, because every vTXO carries a unilateral-exit transaction path its owner can broadcast directly to the base layer. The attack the Ark server can mount is griefing: by refusing to continue rounds, the server forces users into on-chain exits that are expensive in fees during congested periods.

The construction also reveals more information to the server than Lightning reveals to any single channel partner, because the server participates in every user's settlement instead of only a subset. Arkade, the primary production implementation, mitigates part of this exposure by running its signing key inside a Trusted Execution Environment and routing user-to-signer communication through end-to-end encryption, so the operator's infrastructure cannot read individual transaction content even though it coordinates batch settlement.

Arkade's programmable execution environment also supports spending conditions that go beyond what on-chain Bitcoin Script can express, because the off-chain vTXO layer is not constrained by base-layer script limits. In principle this includes zero-knowledge payment schemes where proofs over VTXO outputs replace plaintext

amounts and counterparty identities, collapsing operator visibility toward what Chaumian ecash achieves while preserving unilateral exit. That capability remains theoretical at the time of writing, with no deployed ZK scheme running over Ark in production. Second is the other production-scale implementation at the time of writing; the construction is new enough that its empirical behavior under adversarial operator conditions has not been tested at Lightning's scale.

Spark: Statechain Transfers

Spark is an off-chain Bitcoin protocol built on statechains, extended with FROST threshold signing across multiple Signing Operators (SOs).[24] In the base statechain model, a bitcoin UTXO is controlled by a key split between the current owner and a signing service; transferring the coin off-chain means the current owner and the service jointly reassign the key to the recipient, with the service deleting its prior key share so the previous owner cannot spend. Spark extends this to a distributed set of SOs using FROST and introduces a leaf architecture that allows UTXO values to be split and merged off-chain, removing the whole-coin denomination constraint of earlier statechain designs. The chain records nothing until a leaf exits; on-chain, Spark state appears as a chain of singlesig outputs.

The trust model is 1-of-n at the operator level: as long as one SO honestly deletes its prior key share after a transfer, the previous owner's ability to spend is revoked and the new owner's claim is secure. Even if all SOs later collude or are compelled, they cannot move funds without the current owner's signature, since the user is a required co-signer on every transaction. The attack surface the SOs retain is liveness: a coordinated refusal to co-sign forces the user into a unilateral on-chain exit via a timelocked backup transaction, similar in character to the griefing risk an Ark operator holds.

The privacy ceiling is set by what the SOs see during signing. Every transfer passes through the SO set, which observes the amounts, the sender, and the recipient for each leaf. An outside observer watching the chain sees only funding transactions and exits; the SOs

see the full transfer graph between those events. Splitting the operator role across multiple SOs reduces the single-point-of-compulsion risk (a subpoena or breach must reach a threshold of operators to reconstruct the full history) but no individual SO is blind to the transactions it co-signs. Spark's operator visibility is therefore higher than Lightning's, where each node sees only its adjacent hops, and higher than Ark's batch model, where the operator sees output amounts at settlement but not the internal transfer graph between rounds.

Comparing Lightning, Ark, and Spark

Lightning, Ark, and Spark make different choices along the same trust-and-privacy spectrum. Lightning distributes trust across the channel graph so that each user trusts only their counterparties and can exit unilaterally against any of them; it pays for that distribution in liquidity-management complexity, inbound capacity constraints, and the balance-probing surface described above. Ark concentrates the coordination role in a single server that cannot steal funds but can grief by refusing rounds and sees more of each user's payment flow than any Lightning counterparty sees; it pays for that concentration by eliminating the inbound-liquidity problem that Lightning has not solved for mobile receivers. Spark, anchored to the statechain model, distributes the signing role across a threshold operator set that collectively sees the full transfer graph, in exchange for instant off-chain transfers without locked liquidity or round participation.

The architectural difference between Ark and Spark that carries the most weight for privacy is scriptability. Ark's vTXO layer runs an off-chain execution environment whose spending conditions extend beyond what base-layer Bitcoin Script can express on-chain. Timelocks, hash locks, threshold signatures, and covenant constructs all apply inside an Ark round, but the off-chain context also admits ZK proof systems that have no direct on-chain equivalent: a ZK payment scheme layered over Ark's vTXO model could blind amounts or sender-receiver relationships from the operator entirely, collapsing operator visibility toward what Chaumian ecash achieves

while preserving unilateral exit. Statechain transfers carry no equivalent surface. Each SO must co-sign every state transition, and the signing protocol requires each SO to observe the transaction; there is no script path by which a ZK proof could substitute for the SOs' key contributions. Ark's privacy ceiling is therefore open to future protocol work; Spark's is structurally lower by design.

20.5 Custodial Privacy Layers

Lightning, Ark, and Spark preserve unilateral exit, which is to say each user can reclaim base-layer funds without counterparty cooperation. A second family of privacy layers surrenders that property in exchange for stronger privacy against outside observers, or for confidential amounts, or for receive-offline convenience. The trust model varies across the family, from a single operator to a federation of guardians, and the privacy model varies with it. This section examines the three constructions in order of increasing trust distribution: single-operator mixers, Chaumian mints (single and federated), and the Liquid federated sidechain.

Custodial Mixers as a Privacy Baseline

The crudest route to transaction privacy is to pool funds in a custodian. A centralized Bitcoin mixer accepts deposits from many users, commingles the coins, and returns equivalent amounts from the pool to whichever withdrawal addresses each user specifies. To an outside observer watching the blockchain, the deposits and withdrawals are decorrelated: chain analysis sees that funds entered the operator's cluster and that other funds left it, but not who received what. Against chain-analysis firms and ordinary outside adversaries, the custodian provides real privacy, and this is what Bitcoin mixers have done for more than a decade.

What the custodian does not blind is itself. The mixer operator holds the mapping between deposits and withdrawals by construction, because the operator is the entity moving the coins. When the operator is cooperative, the user's privacy holds against everyone else. When the operator is hostile, subpoenaed, or breached, the

user has no privacy at all, and the operator's internal ledger becomes a surveillance dossier of everyone who ever used the service. The prosecution of Roman Sterlingov for allegedly operating Bitcoin Fog turned on exactly this: the operator saw everything, and the investigation was about reconstructing what the operator had seen. A custodial privacy layer that depends on operator honesty has moved the trust problem, not solved it. Ordinary custodial wallets and exchanges share this structure. Holding bitcoin at a custodian pools on-chain activity behind the custodian's addresses, which provides real privacy against outside chain-analysis observers. What varies is the custodian's own practices: an exchange that requires KYC, logs withdrawals by identity, or sweeps funds through poorly-labeled hot wallets hands that information to any adversary who can compel or breach it. The privacy benefit of custody is real but entirely dependent on the operator's data handling, legal exposure, and on-chain hygiene.

Chaumian Ecash: Blinding the Operator

David Chaum's blind-signature scheme, introduced in 1982, solves the operator problem in exactly the place the custodial mixer cannot.[25] A user prepares a token with a serial number, then blinds the serial number before presenting it to the mint for signature. The mint signs the blinded form without ever seeing the underlying serial. The user unblinds the signed token afterward and holds a valid signature on a serial the mint has no record of. When the user later redeems the token, the mint verifies its own signature on the newly revealed serial but cannot tie the redemption to the specific issuance event.

Applied to Bitcoin, the construction becomes a deposit-and-withdrawal protocol in which the operator genuinely does not know who received which coins. Users deposit bitcoin with a mint and receive blind-signed tokens that function as bearer money warehouse receipts. They transact peer-to-peer in those tokens, or redeem them to bitcoin when they wish to exit. The mint records who deposited and who redeemed but cannot link the two sides of the ledger

together. The operator no longer has the information to betray a user even if compelled.

Cashu and Fedimint: Two Deployments, Same Blinding

Two contemporary projects implement this construction on Bitcoin. Cashu operates as a set of independent single-operator mints speaking a common protocol.[2627] Each mint directly runs a Lightning node, so Lightning is the primary payment rail: users deposit and withdraw via Lightning invoices. The privacy properties follow directly from the Chaumian construction: the operator sees amounts and timings but not the user-to-user flow of tokens, and outside observers of the blockchain see only Lightning channels of the operator. The operational simplicity is also the risk: a single operator can fail, be compromised, disappear, or refuse redemptions, and users of that mint bear the full exposure. Mints are cheap to run, so the practical mitigation is that users hold small balances on any one mint and choose operators whose reputations they can evaluate. Due to the privacy of blind signatures, the mint cannot selectively censor one particular user, only inflate the amount of tokens in circulation, or deny redemption globally and steal all users' funds.

Fedimint implements the same blinding with federated custody.[28] Multiple guardians jointly control the mint's bitcoin reserves using on-chain threshold multisig, so no single guardian can unilaterally steal funds, halt redemptions, or coerce the user. A threshold of guardians must cooperate for any action on the underlying bitcoin. Lightning connectivity comes through independent Lightning Gateways: separate nodes that bridge the federation to the Lightning network and operate outside the federation's trust boundary, so Lightning access does not depend on the federation itself holding channel liquidity. The ecash issuance itself uses a threshold blind signature scheme (TBS) in which each guardian holds only a key share; issuing a blind-signed token requires a threshold of guardians to contribute partial signatures, which means no sub-threshold subset can unilaterally issue or censor tokens. The privacy properties against outside observers are the same as Cashu's, and the blinding of the operator is the same intuition, but the operator now consists of

a federation whose members watch each other. The tradeoff is operational: federations must be set up, coordinated, and maintained, which is more demanding than running a single mint and more demanding than using one.

Liquid: Federated Confidential Transactions

Liquid is a federated Bitcoin sidechain operated by a consortium of exchanges and trading firms under Blockstream's coordination.[29] Its monetary unit is Liquid Bitcoin (L-BTC), a two-way peg to base-layer Bitcoin administered by the federation. Liquid's relevance to this chapter is its Confidential Transactions type, which uses Pedersen commitments and range proofs to hide amounts and asset types on-chain while preserving the federation's ability to audit supply. What Confidential Transactions do not hide is the transaction graph: senders and receivers are visible to any outside observer, and the full payment topology remains public. The privacy gain over base-layer Bitcoin is amount and asset blinding only. Liquid also supports issued assets (tokenized stablecoins and securities tokens) whose amounts ride the same confidentiality machinery.

The trust model is similar to Fedimint's: a federation of known entities jointly controls the peg, any member can observe the transaction graph, and a threshold must cooperate for peg-in or peg-out operations. Privacy against outside observers is weaker than Fedimint's, since the transaction graph is public where Fedimint's ecash flow is not. The federation can, in principle, collude to halt the network or reverse transactions, and the peg depends on the federation honoring redemption. For use cases where amount confidentiality and faster settlement are worth the federation trust (stablecoin issuance, confidential OTC settlement), Liquid has operated continuously since 2018 without incident. For use cases where full transaction-graph privacy or trust-minimization is the requirement, the constructions examined elsewhere in this chapter are the better answer.

Trust Tradeoffs Across the Custodial Spectrum

The custodial layer runs a spectrum along two axes that do not move in lockstep: trust distribution and transaction-graph privacy. Trust distribution increases from single-operator mixers through Cashu's single-operator mints to Liquid's multi-member consortium to Fedimint's threshold-guardian federations. Transaction-graph privacy does not follow the same ordering. Mixers, Cashu, and Fedimint all hide the payment flow from outside observers: chain analysis sees deposits and withdrawals into the operator's cluster but not who paid whom inside it. Liquid hides amounts but leaves the full transaction graph visible on-chain, which means it provides weaker privacy than Cashu or Fedimint against a chain-analysis adversary despite distributing custody across a larger federation.

What is uniform across the spectrum is the absence of unilateral exit. Unlike Lightning, Ark, and Spark, none of these constructions lets a user reclaim base-layer funds without operator cooperation. If the mint or federation goes offline and reserves are lost or seized, the tokens in circulation are unredeemable. They are money substitutes in Mises' sense. Federated models reduce the probability of unilateral operator failure, but the custody risk is present throughout. Sensible practice depends on what amounts, time horizons, and adversary models the user needs to protect against: a user whose primary concern is chain-analysis surveillance benefits most from Cashu or Fedimint; Liquid's transaction-graph visibility and larger, more censorable consortium make it the weakest option on both axes despite its confidential amounts.

20.6 Shielded Client-Side Validation

Moving Validation Off the Chain

All the layers examined so far share a common property: every participant who sees a transaction learns something from it. Base-layer Bitcoin publishes the entire transaction to every full node. CoinJoin and PayJoin publish a transaction whose inputs and outputs are all visible, with privacy coming only from the ambiguity of the join. Lightning hides payment-level activity but publishes

the opening and closing of channels, and its routing graph is public by design. Ecash hides the payment graph from outside observers, and the Chaumian construction blinds the mint to which token was issued to whom. What it cannot remove is the mint itself: every ecash token's validity ultimately rests on the mint's blind signature, so the mint is a required counterparty for every issuance and redemption. A user cannot verify coin validity independently; the mint's continued cooperation and solvency are load-bearing assumptions throughout.

Client-side validation inverts this. Instead of every node validating every transaction, the recipient validates the transactions that move coins to them, and the chain is used only for ordering and to prevent double-spending.[30] The idea originates with Peter Todd's 2013 work on proof-of-publication consensus and was developed most fully in his 2016 paper on state-machine consensus primitives, which introduced the concepts of single-use seals, deterministic transaction expressions, and proof-of-publication as building blocks. A Bitcoin UTXO is already a single-use seal: the outpoint identifies the seal, the script authorizes its closure, and the spending transaction is the proof that it has been closed. Protocols built on top of this observation, most prominently RGB and Taproot Assets, have used Bitcoin as a commitment and ordering substrate for off-chain contract state.

Existing client-side-validation protocols improve privacy over native base-layer transactions, but they do not eliminate the leak. In RGB and similar designs, when a coin is transferred, both the sender and receiver see the full history of the coin, because the proof that validates the current state is the chain of prior state transitions. Coin proofs grow with the length of the transaction history. The on-chain footprint depends on how the commitment is anchored. RGB's recommended method, Tapret, embeds the commitment as a taproot key tweak: the resulting output is indistinguishable from an ordinary taproot spend, so chain analysis cannot identify or cluster RGB transactions from the chain alone. The alternative, Opret, publishes the commitment in an OP_RETURN output, which is visible and clusterable. For Tapret users the binding privacy constraint

is not on-chain visibility but counterparty disclosure: the receiver must see the full prior chain of state transitions to verify the coin, so every transfer leaks history to the immediate recipient.

Shielded CSV: Zero-Knowledge Client-Side Validation

Shielded Client-Side Validation, proposed by Jonas Nick, Liam Eagen, and Robin Linus in a 2025 paper, closes both of those gaps by applying zero-knowledge proofs to client-side validation.[31] The construction uses Proof-Carrying Data, a recursive form of zero-knowledge proof, so that a coin's current holder does not need to see the coin's full history to verify its validity. The sender hands the receiver a succinct proof that the coin in question was produced by a valid chain of Shielded CSV transactions terminating in a legitimate issuance, and the proof's size and verification time are independent of how long that chain is.

What appears on the Bitcoin chain per transaction is 64 bytes of opaque data the paper calls a nullifier, pseudorandom-looking to anyone who does not hold the corresponding receipt. The nullifiers are published as application-layer payload inside ordinary Bitcoin transactions. A Bitcoin full node that is not a Shielded CSV participant treats this payload as raw bytes: it validates the carrier transaction under normal rules, performs no verification on the nullifier itself, and leaves the UTXO set unchanged. Only Shielded CSV participants parse the data, and for them the cost of verifying an entire batch of nullifiers collapses to a single Schnorr verification through the paper's NISSHAC half-aggregation construction. The base chain's role reduces to ordering and proof-of-publication; double-spend prevention is enforced client-side, by Shielded CSV nodes maintaining a key-value store of seen nullifier public keys and rejecting any repeat.

The privacy properties follow from the construction. An outside observer watching the Bitcoin chain sees a stream of 64-byte blobs: amounts, senders, receivers, linkage between transactions, and any off-chain history are all opaque by construction. A receiver validates what they need to accept the coin, and learns nothing about what the sender did with coins they did not transfer. The history that is visible

to counterparties in RGB becomes invisible under Shielded CSV, and the commitment graph that chain-analysis could cluster in RGB collapses to a set of nullifiers that resist clustering by construction.

Maturity and Place in the Layer Stack

Shielded CSV is a research proposal with a prototype implementation, not a deployed system. The paper is recent, the construction depends on recursive proof systems such as folding schemes or recursive STARKs, and the efficiency claims the paper makes rest on a deployment path that does not yet exist.

Its place in the layer stack is nevertheless clear. Lightning remains the right answer for low-latency interactive payments at the base of the stack, and ecash provides instant bearer-style privacy today at the cost of custody. Shielded CSV represents the direction the non-custodial privacy frontier is moving: toward designs where the chain carries commitments it cannot read, the receiver validates proofs whose size does not depend on history, and the information leaked to outside observers collapses to what Bitcoin absolutely requires for consensus. Whether Shielded CSV specifically becomes the deployed standard or another zero-knowledge client-side-validation design takes its place, the architectural shift is the point.

20.7 Privacy Coin Comparators

The privacy layers in this chapter all share a design premise: Bitcoin's base-layer ledger is given, and privacy is bolted on. Two families of alternative cryptocurrency reject that premise and build privacy into the base layer itself. Monero and Zcash are the main comparators. Both have operated continuously for roughly a decade. Neither replaces the Bitcoin analysis in Chapter 18, and neither changes the monetary case for Bitcoin made in Chapter 19. Both earn examination here because they illustrate what privacy-at-the-base-layer buys and what it costs. A third comparator, Ethereum, takes neither the base-layer-privacy route nor the Bitcoin layered route: its base layer is fully transparent, but its programmable contract

environment has produced a set of application-layer privacy overlays whose architectural position clarifies what the layered model requires and where it runs into enforcement limits.

Monero's Construction

Monero encodes three privacy properties directly into the transaction format. Ring signatures hide which input a transaction is spending. Stealth addresses hide which output a recipient controls. Ring Confidential Transactions hide the amount. Each property is a cryptographic primitive with a body of peer-reviewed theory behind it, and the combination produces a ledger on which a casual observer cannot tell who paid whom how much.

A ring signature, originally proposed by Rivest, Shamir, and Tauman in 2001, is a signature scheme in which the signer proves that the private key to one of a specified set of public keys was used, without revealing which one. Monero assembles each spending transaction with a ring of decoys drawn from prior transaction outputs. The network verifies that exactly one output in the ring is being spent, without learning which one. The anonymity set is the ring size, currently sixteen, which means a transaction's true sender is hidden among sixteen candidates from the chain's history. The primitive has a cost. Every ring signature carries the decoys on-chain, and the resulting transaction size is larger than a Bitcoin transaction of comparable semantic content. The per-transaction storage overhead is the visible price of the privacy guarantee.

Ring signatures create an apparent problem: if no one knows which output in the ring is being spent, what stops the same output from being spent twice across different transactions, each time hiding behind a different ring? Monero resolves this with key images. Every spending transaction includes a key image, a value derived deterministically from the private key of the output being spent. The derivation is constructed so that the same output always produces the same key image, regardless of which decoys surround it in a given ring. Nodes maintain a global set of published key images and reject any transaction whose key image has already appeared. A double-spend attempt would reuse the private key of the real

output, producing an identical key image, and would be rejected immediately. The ring can be shuffled arbitrarily; the key image is invariant. The construction lets Monero enforce one-spend-per-output without linking the key image to any specific ring member, because the image is derived from the real key but the ring signature proof does not reveal which ring member supplied it.

A ring of size sixteen is a meaningful improvement over Bitcoin's transparent chain, but it is bounded, and the bound is small compared to zero-knowledge constructions whose anonymity set can approach the entire output population. A single transaction with a ring of sixteen produces one-in-sixteen uncertainty for an adversary with no other information. A sequence of transactions by the same holder, analyzed jointly, reduces that uncertainty. Intersection attacks, in which the adversary observes several rings that all plausibly contain outputs from the same holder, narrow the candidate set multiplicatively across the sequence. The empirical literature on pre-RingCT Monero documented traceability of a substantial fraction of spends through closed-set chain-reaction analysis: if a ring contains only outputs already known to have been spent in other rings plus one unknown output, the unknown must be the real spend. The attack's reach on modern RingCT transactions is much narrower, but the underlying principle means that a ring's effective anonymity depends on the unknowns remaining unknown across the entire transaction graph.

Decoy selection is the operational axis that determines how close the effective anonymity set comes to the nominal sixteen. The first-spend heuristic is the concrete threat: empirical wallet behavior shows that users typically spend recent outputs, so an adversary who guesses that the most recent output in a ring is the real spend succeeds at a rate substantially above chance. Monero's wallet software counters this by drawing decoys from prior outputs according to a gamma distribution weighted toward older outputs, so that recent-output rings contain older decoys that blur the timing signal. The gamma distribution does not close every gap. Decoys drawn from outputs that turn out to have been spent in other rings reduce the effective anonymity; adversaries who contribute a large fraction

of the on-chain output pool through their own transactions shape the decoy distribution available to subsequent users; and wallet implementations that deviate from the recommended selection algorithm produce transactions whose rings are distinguishable from the mainline population and therefore easier to analyze. The Monero development community has responded to each such finding with wallet-level decoy-selection improvements and with hardfork-level ring-size increases, but the underlying constraint remains: a sixteen-member ring cannot provide stronger privacy than the adversary's prior knowledge about which of the sixteen members are plausibly spendable.

A stealth address is a one-time address derived by the sender from the recipient's long-term address. The recipient publishes a single address but receives funds at a different, unlinkable on-chain address for every transaction. The construction uses elliptic-curve Diffie-Hellman between the sender's ephemeral key and the recipient's view key, so that only the recipient can scan the chain and identify incoming payments addressed to them. The chain contains no reusable anchor that a chain-analysis firm could cluster. Stealth addresses defeat the address-reuse heuristic structurally.

Ring Confidential Transactions, deployed on Monero in 2017, extend the ring signature to hide the amount alongside the sender. The construction uses Pedersen commitments to encrypt amounts, and range proofs (currently Bulletproofs and Bulletproofs+) to prove that each hidden amount is a non-negative integer whose sum balances the transaction's inputs and outputs. Without the range proofs, an attacker could create negative amounts and inflate supply; with them, the protocol preserves supply while hiding values. The per-transaction cost of this construction is substantial. A typical Monero transaction occupies roughly ten times the bytes of a Taproot Bitcoin spend, and verifying it costs roughly two orders of magnitude more CPU time on the same hardware, because every node must check the ring signature and the range proof for every input and output. Batching helps Bulletproofs verification within a block but does not help Concise Linkable Spontaneous Anonymous

Group (CLSAG) signature verification, which scales linearly with ring size. The cost is paid by every participant, every block, forever.

Dandelion++, deployed on Monero in 2020, addresses the network-layer attack described in section 20.1. The protocol propagates a new transaction through a randomly chosen stem of relay nodes before broadcasting it to the wider network, so that an observer who captures the first broadcast cannot identify the originating IP address as easily as under naive gossip. The protection is probabilistic and limited against an adversary who controls many vantage points, but it raises the cost of the network-layer attack materially.

Monero's combined architecture makes on-chain clustering much harder than on Bitcoin. The properties hold against casual and moderately well-resourced adversaries. They do not hold against an adversary who can correlate off-chain information, such as an exchange's KYC record tied to a deposit address, nor against an adversary who runs a large fraction of the ring's decoys as their own prior outputs. The arms race between decoy selection and decoy analysis is live.

The protocol-level response in preparation is Full-Chain Membership Proofs. The FCMP and FCMP++ proposals, developed since 2023 primarily by Luke Parker, replace ring signatures with a zero-knowledge proof that the spent output belongs to the set of all spendable outputs on the chain. The anonymity set expands from sixteen to tens of millions, which is the entire output population, and the dependence on wallet-level decoy selection disappears along with the decoy-selection-attack class. The construction uses curve trees expressed over a Helios and Selene curve cycle to commit to the output set succinctly, and the proof size and verification cost remain feasible for consumer-scale wallets. A Monero hardfork enabling FCMP++ is in active preparation at the time of writing. When activated, it will be the largest upgrade to Monero's privacy architecture since RingCT, and it will close the specific class of attacks that exploit the ring's bounded size.[32]

Monero's privacy properties are hardcoded into the transaction format, not expressed through a programmable script, so the pro-

tocol has no general-purpose execution layer. Threshold multisig exists but is an off-chain coordination protocol between cosigners, not an on-chain scriptable condition. HTLCs, which require a hash preimage to be revealed on-chain at settlement, are incompatible with Monero's privacy model; XMR atomic swaps with Bitcoin are built on adaptor signatures instead, avoiding the hash-lock construction entirely. Timelocks exist in a limited form as a coarse block-height mechanism, but conditional timelocks of the kind Bitcoin's CSV and CLTV enable are absent. On-chain ZK verifiers, arbitrary multi-party contracts, and payment channel factories are outside the design space entirely. FCMP++ extends the anonymity set to the full output population but does not change any of this: it is a spending-proof upgrade, not a scripting upgrade. The tradeoff is deliberate. A fixed transaction format enforces privacy uniformly and eliminates the surface area through which a programmable layer could leak metadata. The cost is that Monero cannot host the second-layer payment channels, covenant schemes, or application-layer overlays that give Bitcoin's layered model its extensibility.

Zcash's Construction

Zcash takes a different path. Its base-layer ledger supports both transparent transactions and shielded transactions, with the shielded pool protected by zero-knowledge proofs. A fully shielded transaction reveals neither sender nor receiver nor amount nor any reusable linkage to other transactions; only a nullifier and a new commitment appear on-chain, and the proof accompanying the transaction certifies that the spender owned a valid coin in the shielded pool and has not spent it before. The cryptographic core is a zero-knowledge Succinct Non-Interactive Argument of Knowledge, or zk-SNARK, adapted for the specific statement the transaction must prove.

The protocol has gone through three generations of proof systems. The original Sprout pool, launched in 2016, used a SNARK construction that required a trusted setup ceremony whose compromise would have allowed undetectable counterfeiting. Sapling, activated in 2018, replaced the construction with a more efficient one

and ran a second, multi-party ceremony to reduce the trusted-setup risk. Orchard, activated in the 2022 NU5 upgrade, uses the Halo 2 proof system, which eliminates the trusted setup requirement entirely through a recursive proof construction. The transition across these three generations illustrates an important property of privacy-at-the-base-layer designs: the cryptographic assumptions are not static. A privacy coin must plan for the replacement of its own core primitive, and the transitions must preserve value for existing holders without compromising the privacy of earlier transactions.

Zcash's optional-shielding model has consequences the Monero model does not have. Zcash transparent transactions behave like Bitcoin transactions and can be clustered with the same heuristics. Shielded transactions are private against all of those heuristics. The population that benefits from the privacy is the population that uses the shielded pool, and the shielded pool's usage rate has varied across the protocol's history. When the majority of the economic activity is transparent, an adversary can apply clustering to the transparent edges of shielded-to-transparent transfers and extract information about the shielded pool indirectly. The protocol designers have pursued wallet changes that make shielded transactions the default for new users, and those changes improve the pool's effective anonymity set, but the mixed-mode architecture is a structural feature that a pure-shielded design would not have.

The zk-SNARK construction that makes shielded transactions private also makes them expensive relative to Bitcoin. Sprout shielded transactions were several times larger than comparable Bitcoin transactions and required tens of seconds to prove on consumer hardware; every node on the network had to spend significant CPU verifying each one. The verification cost created a denial-of-service surface. In September 2018, shortly before Sapling activated, an attacker flooded the network with Sprout shielded transactions, bloating the mempool and forcing nodes to expend disproportionate computation per block. The attack slowed synchronization for full nodes materially and exposed a structural property of proof-based privacy: the same computation that certifies a transaction's validity is the mechanism through which an attacker can drain verifier re-

sources. Sapling reduced both transaction size and proving time by over an order of magnitude and made the attack class less viable at Sapling's parameters, but the underlying dynamic remains. A shielded transaction costs more to verify than a transparent one, and any fee market that does not price verification compute accurately leaves room for cheap-to-create, expensive-to-verify spam. Orchard reduced verification cost further and supports batch verification, narrowing the gap, but the privacy premium in bandwidth and computation is a permanent feature of the architecture.[33]

Ethereum's Privacy Architecture

Ethereum's base layer is fully transparent. All transaction data is published to the chain and indexed by block explorers. The address graph is as visible as Bitcoin's, and the same clustering heuristics that chain-analysis firms apply to Bitcoin apply to Ethereum with equal or greater effect, because Ethereum's richer contract interactions create more linking signals, not fewer. The structural reason goes deeper than transparency alone. Bitcoin uses a UTXO model: each received coin is a discrete output at a specific address, and a wallet can generate a fresh address for every incoming payment, breaking the address-reuse heuristic at the wallet layer. Ethereum uses an account model: an address holds a running balance, and every inbound payment accumulates into that same account. Sharing an Ethereum address to receive funds is unavoidable, so every counterparty who sends to it learns the full balance and transaction history attached to that address. ERC-20 token transfers compound this: each token transfer from the same address links the account's ETH activity to its token activity across every contract it has ever called. The account model makes address reuse the base-layer payment mechanism. EIP-5564 stealth addresses offer a mitigation (a sender derives a one-time address from a recipient's public key so the recipient alone can identify incoming funds) but adoption remains marginal and the base-layer account structure is unchanged for users who do not generate stealth addresses for every inbound payment.

The "ZK rollups" that have made Ethereum a prominent topic in zero-knowledge proof discussions do not change this. ZK rollups (zkSync Era, StarkNet, Scroll, Polygon zkEVM) use cryptographic proofs to certify that the operator computed the state transition correctly; they publish all transaction data to Ethereum as calldata so that anyone can reconstruct the full L2 account history. The zero-knowledge proof answers whether the computation was correct, not who paid whom. The mechanics are covered in Chapter 15; the architectural point here is that a ZK rollup provides validity guarantees on a transparent ledger, not privacy on an opaque one. An outside observer who indexes the calldata knows every sender, receiver, and amount on every ZK rollup that follows this architecture.

Privacy on Ethereum is an application-layer property, achieved by smart contracts that implement shielded pools. The two deployed systems of note are Railgun and Aztec. Railgun operates as a set of immutable smart contracts on Ethereum mainnet, accepting deposits of ETH and ERC-20 tokens into a shielded pool where internal transfers use zero-knowledge proofs. An observer on-chain sees tokens enter and exit the Railgun contracts; internal transfers between shielded addresses are not visible, and deposits cannot be linked to specific withdrawals. The anonymity set for any given token is limited to that token's pool size, which creates the same weak-pool problem Zcash's optional-shielding model has: privacy depends on how many other users are shielding the same token. Aztec launched its Alpha mainnet in March 2026 as a private-by-default Layer 2. Its architecture is closer to Zcash's shielded model than to a ZK rollup: private state is represented as encrypted UTXO-style notes, private functions execute client-side, and only nullifiers and encrypted commitments appear on-chain. The Alpha network carried a disclosed critical vulnerability at launch, with a fix planned for a later release; the project positioned it as early-stage infrastructure with known security caveats. Aztec is the closest Ethereum analogue to a Zcash shielded pool, available as an opt-in L2 on top of an otherwise transparent base. Railgun's explicit positioning around association set compliance, following the framework Vitalik Buterin and co-authors described in a 2023 paper on Privacy Pools, reflects

the enforcement environment that Tornado Cash's 2022 OFAC designation established. Chapter 13 covers the legal arc and the *Van Loon* ruling that vacated the protocol designation.[34]

The Comparative Analysis

The privacy guarantees Monero and Zcash provide are strong, and Ethereum's shielded pool systems provide real unlinkability within their pools. None of the core constructions in these systems have been broken. The architectural difference that matters is the breadth of assumptions and the size of the resistance surface, not privacy strength alone.

Bitcoin's base-layer consensus relies on well-understood primitives: elliptic-curve signatures on secp256k1, the SHA-256 hash function, and the proof-of-work security model built on them. The privacy layers above Bitcoin add primitives incrementally. CoinJoin adds only coordination; it introduces no new cryptographic assumption. PayJoin adds only a specific transaction-construction pattern. Lightning adds HTLC payments and the specific revocation-key dance, both built from the base primitives. Ark adds off-chain vTXOs with a shared UTXO construction; Spark adds off-chain balances via FROST threshold signing across statechain operators, each introducing no new on-chain consensus assumption. Chaumian ecash adds blind signatures, which rely on the same elliptic-curve assumptions. Shielded client-side validation introduces zero-knowledge proofs; those proofs rely on additional assumptions.

Monero and Zcash adopt the heavier cryptographic machinery as a condition of participation in the protocol at all. Every Monero transaction relies on ring signatures and Pedersen commitments and range proofs, and if any of those constructions failed, every transaction would be affected. Every Zcash shielded transaction relies on the zk-SNARK proof system in use at the time the transaction was made, and a failure in the proof system would be catastrophic. The Halo 2 transition reduced this risk by removing the trusted-setup assumption; it did not eliminate the dependence on the broader SNARK construction.

The cost asymmetry argument the book develops in Chapter 10 therefore runs differently for these protocols. Bitcoin keeps its base-layer resistance surface small and moves privacy into composable higher layers where failures are contained. Monero and Zcash build privacy into the base layer and accept a larger resistance surface in exchange. Both approaches are coherent. Which approach is more persuasive depends on whether the privacy guarantee at the protocol level matters more than the conservatism of the assumption set.

A second architectural difference affects what can be built on top of these base layers. Bitcoin Script supports multisig custody, timelocks, hash locks, HTLCs, Lightning's revocation-key construction, discreet log contracts, and the script commitments under which client-side validation operates. Monero's transaction format has no user-facing scripting, and Zcash's scripting remains narrower than Bitcoin's even after what Sapling and Orchard expose. The absence is not an oversight: expressive on-chain conditions reveal the structure the privacy layer is meant to hide, and the two protocols have consistently resolved the conflict in favor of privacy. The consequence is that Lightning-class payment channels, HTLC-routed cross-chain atomic swaps, and non-custodial escrow of the kind Bitcoin supports do not exist natively on Monero or Zcash. Privacy-preserving scaling on these chains happens at the base layer, with the verification-cost consequences already described, instead of in the composable higher layers that Bitcoin's limited-but-usable scripting enables. A third approach, distinct from both base-layer anonymity and Bitcoin's layered scripting, is the zero-knowledge smart-contract model that Amir Taaki's DarkFi has been developing: every contract is expressed as a ZK circuit whose proof is published on the chain, so arbitrary programmable logic runs without revealing the inputs or outputs to anyone verifying the proof. That model offers expressive contracts and privacy simultaneously, at the cost of a much heavier cryptographic assumption set and an early-stage implementation whose performance envelope is still being established.[35]

Regulators and exchanges have been delisting privacy coins for most of a decade, from the first Japanese withdrawals after the 2018 Coincheck hack through each successive wave of Travel Rule and MiCA enforcement. The underlying cause is structural: base-layer privacy conflicts with the Travel Rule and with the KYC edge the chain-analysis industry uses to anchor clusters, so any exchange that lists a privacy coin inherits a compliance problem that is cheaper to resolve by delisting. Bitcoin does not provoke the same conflict at the base layer, and privacy on Bitcoin is an achievement of higher layers built around a listed base currency. The same compliance pressure still reaches Bitcoin through a narrower channel. Chain-analysis firms flag coinjoin outputs as high risk, and major exchanges have rejected deposits traced to Wasabi or JoinMarket participations. Users who receive such outputs have had accounts frozen or funds held pending source-of-funds review, so Bitcoin users exercising the higher-layer privacy tools pay their costs at the on-ramp and off-ramp instead of at the protocol. The difference across the three regimes is architectural: Monero and Zcash fight for their base layer to be listed at all; Bitcoin fights over which higher-layer constructions downstream custodians will accept; Ethereum's application-layer privacy contracts occupy a middle position, subject to direct contract-address sanctions that Bitcoin's protocol-layer tools cannot receive, but not carrying the delisting pressure that base-layer privacy coins face.[36]

20.8 The Quantum Horizon for Bitcoin

Bitcoin's spending authorizations rest on the discrete-logarithm problem in the secp256k1 elliptic-curve group. Shor's algorithm solves that problem in polynomial time on a sufficiently powerful quantum computer, so any coin whose public key is visible on-chain becomes spendable by whoever runs the attack first. The threat is specific to the signature scheme; SHA-256 used for proof-of-work stays hard under Grover's quadratic speedup and is not at risk. The vulnerable population divides into two groups: coins at Pay-to-Public-Key addresses from Bitcoin's earliest years (a small transac-

tion count but large value, including most of the supply attributed to Satoshi), and coins at reused or previously-spent Pay-to-Public-Key-Hash addresses whose public keys are already on-chain. Credible published estimates put roughly a quarter of the circulating supply in one of these categories. Coins held at addresses that have never revealed their public key are safe until the moment of spending, and spending itself becomes the window of attack.

The upgrade path is two steps, each a separate proposal. BIP-360 defines Pay-to-Merkle-Root, a SegWit version-2 output type with `bc1z` addresses. P2MR removes Taproot's key-path spend, leaving only the script-tree path, which closes the long-exposure vulnerability: an attacker who harvests a public key from an old P2MR output gains nothing, because the key path no longer exists. A second proposal, not yet written, will layer post-quantum signature opcodes onto the P2MR script tree to close the short-exposure window during mempool broadcast. Neither step is activated at the time of writing. The harder question is not cryptographic but political and economic: what to do with the roughly quarter of supply already at risk. Proposals range from mandatory migration within a deadline to permanent confiscation of unmoved coins at the moment a quantum attack becomes feasible. Every option distributes losses differently across holders, and the community has not resolved the question.

Signature size is the economic cost of the migration, and it runs directly into every layer above the base. A Schnorr signature is 64 bytes. FALCON, the compact lattice-based scheme NIST selected for ongoing standardization, produces 666 bytes at the 128-bit security level, a factor of ten over Schnorr. ML-DSA (FIPS 204, finalized August 2024) runs 2,420 to 4,595 bytes depending on security level. SLH-DSA (FIPS 205, finalized August 2024), the standardized hash-based family, runs 7,856 bytes in its smallest parameter set to nearly 50,000 bytes in its fastest. Kudinov and Nick of Blockstream Research showed in December 2025 that SPHINCS+ variants tuned to Bitcoin's requirements, reducing signatures per key and applying SPHINCS+C and PORS+FP optimizations, achieve meaningfully smaller signatures than standardized SLH-DSA while relying only

on hash assumptions already present in Bitcoin's security model.[37] Blockstream Research has taken this further with SHRINCS, a hybrid scheme pairing a stateful XMSS-derived path (approximately 1,092 bytes for a first spend, growing with subsequent spends from the same key) with a stateless fallback at 4,396 bytes, deployed experimentally on the Liquid sidechain. The direction of research is clear: Bitcoin-specific hash-based designs converge toward signatures in the low thousands of bytes, against Schnorr's 64, with verification costs close enough to Schnorr to remain node-accessible. Every privacy layer above the base inherits this cost. A Lightning channel open, a CoinJoin input, a Shielded CSV authorization: each carries the new signature, and the fee-market arithmetic of the post-quantum protocol is an open question the community must resolve before the threat arrives.[38]

Chapter Summary

Bitcoin's base layer is transparent by design, and a surveillance industry has built itself around address reuse, the common-input-ownership heuristic, change detection, network-layer observation, and KYC-anchored cluster identification. Privacy on Bitcoin is an architectural achievement built atop that base, with each layer introducing distinct tradeoffs along two axes: on-chain versus off-chain, and non-custodial versus custodial. Script primitives (multisig, timelocks, hash locks, HTLCs) enable layered construction without trusted intermediaries. CoinJoin and PayJoin improve on-chain privacy without changing consensus rules. CoinSwap severs the link between spent and received coins across separate ordinary-looking transactions with no on-chain mixing fingerprint; the construction is mainly theoretical on Bitcoin today, with Lightning serving as its off-chain production analogue. Lightning, Ark, and Spark move payments off-chain while preserving unilateral exit, at differing costs in operator visibility, liquidity management, and scripting capability. Custodial layers surrender unilateral exit for stronger privacy along a spectrum of increasing trust distribution: mixers conceal flows from outside observers but not from the operator, Chaumian

ecash (Cashu and Fedimint) blinds the operator cryptographically at the cost of custody, and the Liquid federated sidechain distributes custody across a consortium while hiding amounts but leaving the transaction graph visible. Shielded client-side validation pushes the frontier further by publishing only opaque commitments the chain cannot read, in exchange for current research-stage maturity.

Monero and Zcash build privacy into the base layer, accepting a larger cryptographic assumption set and less programmability in exchange for default privacy. Monero's ring signatures hide the real input among sixteen decoys, key images prevent double-spend without revealing the real spender, stealth addresses break address-reuse clustering, and Ring Confidential Transactions hide amounts at the cost of transaction sizes roughly ten times a comparable Bitcoin spend. Decoy-selection attacks narrow the effective anonymity set below sixteen; FCMP++ replaces rings with a zero-knowledge proof over the entire output population, eliminating the attack class entirely, with a hardfork in preparation. Because privacy is hardcoded into the transaction format, Monero has no user-facing scripting: Lightning-class channels, HTLCs, and covenant schemes are absent by design. Zcash's zk-SNARKs shield sender, receiver, and amount together across three proof-system generations (Sprout 2016, Sapling 2018, Orchard with Halo 2 2022), with Halo 2 eliminating the trusted-setup assumption. The optional-shielding model means transparent and shielded transactions coexist: when most economic activity is transparent, chain analysis can extract information about the shielded pool from the edges of shielded-to-transparent transfers. Shielded transactions cost more to verify than transparent ones, a structural property that created a denial-of-service surface in 2018 when Sprout spam bloated the mempool; Sapling and Orchard reduced the gap without closing it.

Ethereum's base layer is as transparent as Bitcoin's, with an account model that makes address reuse the default payment mechanism: every counterparty who sends to an address learns its full balance and history. ZK rollups publish all transaction data to Ethereum as calldata; the zero-knowledge proof certifies correct computation on a transparent ledger, not privacy on an opaque one.

Privacy on Ethereum is an application-layer property: Railgun implements shielded pools via immutable smart contracts on mainnet, and Aztec launched a private-by-default Alpha mainnet in March 2026 with a disclosed critical vulnerability and a fix planned for a later release. Both face a distinct enforcement channel: OFAC sanctioned Tornado Cash's contract addresses directly in 2022, a form of pressure structurally unavailable against Bitcoin's coordination-only privacy protocols. A decade of regulatory waves (Japan 2018, Bittrex 2021, MiCA-driven 2023–2024) has produced successive privacy-coin delistings because base-layer privacy conflicts with the Travel Rule at the protocol. Compliance pressure reaches Bitcoin privacy through a narrower channel: chain-analysis firms flag coinjoin outputs as high risk, and several exchanges have rejected deposits traced to Wasabi or JoinMarket, so users exercising higher-layer privacy tools pay their costs at the on-ramp and off-ramp.

The post-quantum horizon bounds every layer above Bitcoin's signature scheme. Shor's algorithm breaks secp256k1 on a sufficient quantum computer; BIP-360's Pay-to-Merkle-Root output type removes Taproot's key-path spend as a first step toward migration, with post-quantum signature opcodes to follow. ML-DSA (FIPS 204) runs 2,420 to 4,595 bytes per signature depending on security level; SLH-DSA (FIPS 205) runs 7,856 to 49,856 bytes across its parameter sets, with a size-speed tradeoff between small and fast variants; FALCON, whose FIPS standardization was still in progress at the time of writing, runs approximately 666 to 1,280 bytes. Kudinov and Nick (2025) show that SPHINCS+ variants optimized for Bitcoin achieve meaningful size reductions over standardized SLH-DSA while retaining hash-only security assumptions. Every figure is against Schnorr's 64 bytes, so the transition imposes fee pressure across every privacy layer that carries a spending authorization. No single layer provides complete financial privacy; each tool addresses specific metadata problems while leaving others unresolved, and operational discipline across layers, examined in Chapter 22, remains essential. The anonymous-communication foundations beneath these monetary layers (VPNs, Tor, mixnets) belong to Chapter 17.

Notes

[1]Working tools for readers who want to apply what this chapter covers: Sparrow Wallet (sparrowwallet.com) is the strongest desktop Bitcoin wallet for coin control, label hygiene, and PayJoin/CoinJoin integration; JoinMarket (github.com/JoinMarket-Org/joinmarket-clientserver) remains an actively maintained CoinJoin implementation; Wasabi Wallet (wasabiwallet.io) ships WabiSabi-style coinjoins; Mempool.space, OXT (oxt.me), and KYCP (kycp.org) expose the same analytics that chain-analysis firms use, which is useful for learning what your transactions look like to them. For Lightning self-custody, Phoenix (phoenix.acinq.co), Zeus (zeusln.com), and Breez (breez.technology) ship mobile clients with different trust profiles; Core Lightning (corelightning.org) and LND (lightning.engineering) are the primary node implementations. On reading: the Bitcoin Wiki's "Privacy" page (en.bitcoin.it/wiki/Privacy) is the best-maintained encyclopedic reference, maintained by practitioners; Andy Greenberg, *Tracers in the Dark* (2022) is the definitive journalistic account of the surveillance industry this chapter describes; Bitcoin Optech (bitcoinops.org) publishes the best ongoing coverage of protocol-level developments. For Monero as the benchmark alternative privacy design, Sarang Noether and Brandon Goodell's Monero Research Lab reports (github.com/monero-project/research-lab) are worth reading.

[2]Pieter Wuille, "BIP32: Hierarchical Deterministic Wallets" (2012); Marek Palatinus and Pavol Rusnák, "BIP44: Multi-Account Hierarchy for Deterministic Wallets" (2014); BIP84 and BIP86 extend the scheme for SegWit and Taproot outputs respectively. For a practitioner overview of reuse patterns and their prevalence, see the Bitcoin Wiki, "Address reuse," `https://en.bitcoin.it/wiki/Address_reuse`.

[3]Sarah Meiklejohn, Marjori Pomarole, Grant Jordan, Kirill Levchenko, Damon McCoy, Geoffrey M. Voelker, and Stefan Savage, "A Fistful of Bitcoins: Characterizing Payments Among Men with No Names," *ACM Internet Measurement Conference* (2013). The heuristic is also noted in Satoshi Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System* (2008), section 10. For the subsequent demonstration that reuse plus co-spending creates "super-clusters" linking many identities, see Martin Harrigan and Christoph Fretter, "The Unreasonable Effectiveness of Address Clustering," arXiv:1605.06369 (2016).

[4]The change heuristics described here are catalogued on the Bitcoin Wiki, "Privacy," `https://en.bitcoin.it/wiki/Privacy`, drawing on Meiklejohn et al. (2013), Elli Androulaki et al., "Evaluating User Privacy in Bitcoin," *Financial Cryptography* (2013), and Jonas Nick, "Data-Driven De-Anonymization in Bitcoin," MSc thesis, ETH Zürich (2015). For a practitioner treatment of wallet fingerprinting, see 0xB10C's ongoing work at `https://transactionfee.info`.

[5]0xB10C, "LinkingLion: An Entity Linking Bitcoin Transactions to IPs?", observations post, `https://b10c.me/` (2023), with follow-up update in 2024. For the earlier academic work on network-layer deanonymization, see Alex

Biryukov, Dmitry Khovratovich, and Ivan Pustogarov, "Deanonymisation of Clients in Bitcoin P2P Network," *ACM CCS* (2014); and Philip Koshy, Diana Koshy, and Patrick McDaniel, "An Analysis of Anonymity in Bitcoin Using P2P Network Traffic," *Financial Cryptography* (2014). For Dandelion++: Giulia Fanti et al., "Dandelion++: Lightweight Cryptocurrency Networking with Formal Anonymity Guarantees," *SIGMETRICS* (2018).

[6]Financial Action Task Force (FATF), "Guidance for a Risk-Based Approach to Virtual Assets and Virtual Asset Service Providers" (June 2019), `https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Guidance-rba-virtual-assets-2019.html`. The Travel Rule requirement (FATF Recommendation 16 applied to VASPs) mandates that originator and beneficiary identifying information travel with transfers above USD/EUR 1,000. FATF, "Updated Guidance for a Risk-Based Approach: Virtual Assets and VASPs" (October 2021), `https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Guidance-rba-virtual-assets-2021.html`. National rollout varied: FinCEN applied existing BSA rules; the EU's Transfer of Funds Regulation (TFR) extended Travel Rule to all crypto transfers regardless of amount from December 2024 under MiCA's companion regulation. For a compliance-industry survey of national implementation status, see CipherTrace, "FATF Travel Rule Global Implementation" (2023), `https://ciphertrace.com/`.

[7]Chainalysis, founded 2014, `https://www.chainalysis.com/`. Elliptic, founded 2013, `https://www.elliptic.co/`. TRM Labs, founded 2018, `https://www.trmlabs.com/`. Mastercard's CipherTrace, acquired by Mastercard 2021, `https://ciphertrace.com/`. Each vendor maintains a proprietary labeled-address database assembled from exchange partnerships, subpoenas, deposit-address harvesting, and open-source intelligence; the commercial contracts through which exchange KYC data flows to these vendors are generally not public. For independent academic analysis of how the labeling ecosystem works, see Kimberly Easton, Malte Möser, and Tyler Moore, "An Analysis of the Cryptocurrency Industry's Transaction Monitoring Ecosystem," *Financial Cryptography* (2022).

[8]United States v. Roman Sterlingov, No. 21-cr-00399 (D.D.C.). Sterlingov was arrested in April 2021 and convicted by jury in March 2024 on money-laundering conspiracy charges arising from his alleged operation of Bitcoin Fog, a centralized mixing service active from 2011 to 2021. The trial became a high-profile test case for chain-analysis attribution: the government relied heavily on Chainalysis Reactor testimony, and defense expert Elizabeth Bisbee challenged the methodology's reproducibility and peer-review status. The court admitted the evidence; the verdict is under appeal. For the trial record and expert testimony transcripts, see the PACER docket for 1:21-cr-00399. Journalistic coverage: Andy Greenberg, *Wired*, March 12, 2024, "The Controversial Evidence That Convicted Bitcoin Fog's Alleged Operator."

[9]For full-node wallet operation, Bitcoin Core at `https://bitcoincore.org/` is the reference implementation; current storage

footprint is approximately 700 GB with ∼80–100 GB annual growth, initial sync 6–24 hours on consumer SSD. The canonical self-hosting guide is Jameson Lopp's resource collection at https://github.com/jlopp/civil-disobedience. Sparrow Wallet (https://sparrowwallet.com/) and Specter Desktop (https://specter.solutions/) are the two most common desktop wallets that pair with a local Bitcoin Core node. On the general architecture, see the Bitcoin Wiki "Privacy" page, https://en.bitcoin.it/wiki/Privacy, particularly §4.3 and §5.3 on wallet history retrieval.

[10]BIP 157 (Olaoluwa Osuntokun, Alex Akselrod, and Jim Posen) and BIP 158 (Olaoluwa Osuntokun and Alex Akselrod), "Client Side Block Filtering" (2017), https://github.com/bitcoin/bips/blob/master/bip-0157.mediawiki and https://github.com/bitcoin/bips/blob/master/bip-0158.mediawiki. BIP 37 (Mike Hearn and Matt Corallo), "Connection Bloom filtering" (2012), https://github.com/bitcoin/bips/blob/master/bip-0037.mediawiki, disabled by default in Bitcoin Core via Matt Corallo's PR #16152 (merged 2019-07-19, shipped in 0.19.0). The privacy critique motivating the shift: Arthur Gervais, Ghassan O. Karame, Damian Gruber, and Srdjan Capkun, "On the Privacy Provisions of Bloom Filters in Lightweight Bitcoin Clients," *ACSAC '14* (2014), https://eprint.iacr.org/2014/763; Jonas Nick, "Privacy in BitcoinJ" (February 12, 2015), https://jonasnick.github.io/blog/2015/02/12/privacy-in-bitcoinj/, empirically validates the attack. Bitcoin Development Kit (BDK) exposes the BIP 157/158 filter primitives to wallet developers at https://bitcoindevkit.org/, and the filter index is supported server-side by btcd (https://github.com/btcsuite/btcd) and by Bitcoin Core's opt-in blockfilterindex. Wasabi Wallet (https://wasabiwallet.io/) pioneered consumer-wallet client-side filtering with Wasabi 1.0 (August 2018) using a Wasabi-specific filter served from its own backend; Wasabi 2.6.0 (May 2025) added support for standard BIP 158 filters and direct Bitcoin Core node connection, release notes at https://github.com/WalletWasabi/WalletWasabi/releases/tag/v2.6.0. For self-hosted Electrum-protocol servers, ElectrumX (https://github.com/spesmilo/electrumx), electrs (Roman Zeyde, https://github.com/romanz/electrs), and Fulcrum (Calin Culianu, https://github.com/cculianu/Fulcrum) are the three standard implementations.

[11]Stratum V2 specification at https://stratumprotocol.org/, with the Noise-framework encrypted transport and optional Job Declaration extension documented in the core specification. Stratum V2 production deployments: Braiins Pool (https://braiins.com/pool), DEMAND Pool (https://dmnd.work/), Public Pool (https://web.public-pool.io/). Reference implementation at https://github.com/stratum-mining/stratum. OCEAN Pool runs the

parallel DATUM (Decentralized Alternative Templates for Universal Mining) protocol instead of Stratum V2; DATUM specification at https://ocean.xyz/docs/datum and source at https://github.com/OCEAN-xyz/datum_gateway. On Bitcoin's transaction-broadcast privacy via fresh-Tor-circuit-per-broadcast, see Wasabi Wallet's documentation at https://docs.wasabiwallet.io/ and Bitcoin Core's -onlynet=onion plus per-peer circuit isolation as described in the Tor configuration notes at https://github.com/bitcoin/bitcoin/blob/master/doc/tor.md. For the Dandelion++ denial-of-service analysis that kept the proposal out of Bitcoin Core, see the Bitcoin Core PR discussion on the original Dandelion++ draft, referenced at https://github.com/bitcoin/bitcoin/issues/10054.

[12]MuSig2: Jonas Nick, Tim Ruffing, and Yannick Seurin, "MuSig2: Simple Two-Round Schnorr Multi-Signatures," *CRYPTO 2021*, IACR ePrint 2020/1261, https://eprint.iacr.org/2020/1261. Standardized as BIP 327: Jonas Nick, Tim Ruffing, Yannick Seurin, and Elliott Jin, "MuSig2: Schnorr Multi-Signature Protocol," https://github.com/bitcoin/bips/blob/master/bip-0327.mediawiki. FROST (Flexible Round-Optimized Schnorr Threshold signatures): Chelsea Komlo and Ian Goldberg, "FROST: Flexible Round-Optimized Schnorr Threshold Signatures," *Selected Areas in Cryptography* (2020), IACR ePrint 2020/852, https://eprint.iacr.org/2020/852. FROST is the basis of Spark's signing protocol and of several Bitcoin wallet threshold-custody implementations; it is also the signature scheme underlying FCMP++ leaf proofs in Monero. For threshold ECDSA as the pre-Schnorr alternative, see Rosario Gennaro and Steven Goldfeder, "Fast Multiparty Threshold ECDSA with Fast Trustless Setup," *ACM CCS 2018*.

[13]BIP 341 (Taproot), Pieter Wuille, Jonas Nick, and Anthony Towns, "Taproot: SegWit version 1 spending rules" (2020), https://github.com/bitcoin/bips/blob/master/bip-0341.mediawiki. Activated at block 709,632, November 14, 2021. BIP 342 (Tapscript), same authors, https://github.com/bitcoin/bips/blob/master/bip-0342.mediawiki. BIP 340 (Schnorr signatures), Pieter Wuille, Jonas Nick, and Tim Ruffing, https://github.com/bitcoin/bips/blob/master/bip-0340.mediawiki. The privacy analysis of the key-path vs. script-path spend is in BIP 341 §Motivation; the point that unused script branches remain unpublished is central to the design. For the activation history (Speedy Trial, signaling, lock-in), see Bitcoin Optech Newsletter #139 and #157, https://bitcoinops.org/.

[14]Gregory Maxwell, "CoinJoin: Bitcoin Privacy for the Real World," Bitcoin Forum post (2013). On WabiSabi, see Yuval Kogman, Max Hillebrand, Alex Wied, and Aviv Zohar, "WabiSabi: Centrally Coordinated CoinJoins with Variable Amounts," IACR ePrint 2021/206, https://eprint.iacr.org/2021/206; the WabiSabi protocol notes at https:

//github.com/WalletWasabi/WabiSabi/blob/master/protocol.md; and Bitcoin Optech Newsletter #102, "WabiSabi coordinated coinjoins with arbitrary output values," https://bitcoinops.org/en/newsletters/2020/06/17/#wabisabi-coordinated-coinjoins-with-arbitrary-output-values.

[15]JoinMarket, available at https://github.com/JoinMarket-Org/joinmarket-clientserver. Originally developed by Chris Belcher and Adam Gibson (waxwing); Adam Gibson's design essays, including "JoinMarket: Rethinking Coinjoin" (2017), explain the market-maker/taker model and the privacy rationale. JoinMarket's Orderbook (https://joinmarket.me/ob/) shows active maker liquidity in real time. For the PayJoin-inside-JoinMarket extension (P2EP), see Adam Gibson's blog at https://reyify.com/blog/. The protocol's resistance to coordinator takedown stems from the fact that each taker coordinates its own round directly with chosen makers over an encrypted messaging layer (currently IRC and Tor hidden services), so there is no single coordinator server to subpoena or block.

[16]The zkSNACKs coordinator (wasabiwallet.io) announced on June 1, 2024 that it would stop coordinating CoinJoin transactions for U.S.-linked UTXOs following Department of Justice actions against Samourai Wallet (April 2024). The coordinator fully ceased operation on June 1, 2024. On the Samourai Wallet prosecution: United States v. Keonne Rodriguez and William Lonergan Hill, No. 24-cr-00242 (S.D.N.Y.), indictment unsealed April 24, 2024, charging money laundering conspiracy and operating an unlicensed money transmitting business through Samourai's Whirlpool CoinJoin and Ricochet features. The alternative coordinators that emerged include WabiSabi-compatible coordinators operated by BTCPay Server (https://btcpayserver.org/) and independent operators discoverable through Wasabi's coordinator-list mechanism; see https://wasabiwallet.io/swagger/index.html for the coordinator API. The WabiSabi protocol specification is open at https://github.com/WalletWasabi/WabiSabi.

[17]The unnecessary-input heuristic was applied specifically to PayJoin and P2EP detection by Adam Gibson (waxwing) and LaurentMT in a December 2018 GitHub gist that defined the UIH1 and UIH2 variants and provided the first empirical baseline against mainnet history; see https://gist.github.com/AdamISZ/4551b947789d3216bacfcb7af25e029e. Subsequent refinement to subset-sum form appeared in the same thread and was incorporated into the BTCPay PayJoin implementation through https://github.com/btcpayserver/btcpayserver/pull/4600. The academic treatment is Simin Ghesmati, Walid Fdhila, and Edgar Weippl, "Unnecessary Input Heuristics & PayJoin Transactions," IACR Cryptology ePrint Archive, Paper 2022/589 (2022), https://eprint.iacr.org/2022/589. The general unnecessary-input idea predates these and is curated in Chris Belcher's Bitcoin Wiki "Privacy" article at https://en.bitcoin.it/wiki/Privacy under the change-detection heuristics. BIP 78's own treatment of impacted heuristics, mixed input types, and fingerprint defenses is Nicolas Dorier, "BIP 78:

A Simple Payjoin Proposal," https://github.com/bitcoin/bips/blob/master/bip-0078.mediawiki.

[18]BIP 77 ("Async Payjoin") is Dan Gould and Yuval Kogman, https://github.com/bitcoin/bips/blob/master/bip-0077.md, first merged as a draft in May 2025. The reference implementation lives at https://github.com/payjoin/rust-payjoin, with the OHTTP relay component at https://github.com/payjoin/ohttp-relay and the production directory deployment at https://payjo.in/. Adoption status as of writing tracks the curated list at https://en.bitcoin.it/wiki/PayJoin_adoption: BTCPay Server, JoinMarket, Bull Bitcoin Mobile, and Cake Wallet ship both sender and receiver support; Wasabi, Sparrow, BlueWallet, Mutiny, and Sideshift ship send-only; Bitcoin Core has no native integration (tracking issue at https://github.com/bitcoin/bitcoin/issues/19148). The BIP 77 motivation section's own statement that "BIP 78's requirements for Payjoin Version 1 have proven to be an obstacle to adoption" is the most direct primary-source acknowledgment of the v1 infrastructure barrier.

[19]Tier Nolan, "Alt chains and atomic transfers," Bitcoin Forum, May 2013, https://bitcointalk.org/index.php?topic=193281.msg2003765. This post is the earliest documented proposal for atomic cross-chain swaps using hash-timelock contracts. The same-chain CoinSwap application was developed subsequently; for the earliest formal treatment of CoinSwap as an on-chain privacy tool, see Greg Maxwell's discussion in the CoinJoin thread (2013) and Chris Belcher's "CoinSwap: Transaction graph disjoint trustless trading" proposal (2020), https://gist.github.com/chris-belcher/9144bd57a91c194e332fb5ca371d0964.

[20]Joseph Poon and Thaddeus Dryja, "The Bitcoin Lightning Network: Scalable Off-Chain Instant Payments" (2016).

[21]On Lightning privacy limitations, see Sergei Tikhomirov et al., "An Empirical Analysis of Privacy in the Lightning Network," *Financial Cryptography* (2021); Nishanth Chandran et al., "On the Difficulty of Hiding the Balance of Lightning Network Channels," *IACR ePrint* (2019). For a practitioner-oriented treatment of routing analysis, channel coinjoins, blinded paths, and trampoline routing as Lightning privacy improvements, see ben Carman, Evan Kaloudis, Max Hillebrand, Paul Miller, and Tony Giorgio, "Lightning Privacy Research," available at https://lightningprivacy.com and at https://github.com/BitcoinDevShop/lightning-privacy-research.

[22]Matteo Romiti et al., "Cross-Layer Deanonymization Methods in the Lightning Protocol," *Financial Cryptography* (2021). The research showed clustering heuristics linking Lightning nodes to Bitcoin addresses through shared naming, hosting information, and on-chain interactions.

[23]Burak Keçeli, "Ark: A Privacy-Preserving Layer 2 for Bitcoin" (2023), specification at https://arkdev.info/, with additional documentation at https://ark-protocol.org/. Arkade (https://arkadeos.com/) is the primary production implementation, operated by Ark Labs; its stack is at

https://github.com/arkade-os. Second (https://second.tech/) is the other production implementation, operating an Ark server on Bitcoin mainnet with a parallel signet deployment; its wallet implementation (bark) is at https://gitlab.com/ark-bitcoin/bark. For comparison with the Lightning liquidity model the construction addresses, see Bastien Teinturier, "Ark and the Future of Bitcoin L2s," *Bitcoin Optech* newsletter coverage, 2023–2025, at https://bitcoinops.org/.

[24]Spark protocol: Buildonspark, "Spark: A Bitcoin Statechain Protocol," specification at https://docs.spark.money/ and https://github.com/buildonspark/spark. Spark extends the statechain model originated by Ruben Somsen: Ruben Somsen, "Statechains: Non-custodial Off-chain Bitcoin Transfer," Bitcoin Optech Newsletter #91 (2020), and the full proposal at https://github.com/RubenSomsen/rubensomsen.github.io/blob/master/img/statechains.pdf. The FROST threshold extension across multiple Signing Operators is Spark's key departure from the original single-operator statechain design. For Somsen's original statechain proposal, see also the bitcoin-dev mailing list thread from March 2020.

[25]David Chaum, "Blind Signatures for Untraceable Payments," *Advances in Cryptology: CRYPTO '82* (1983): 199-203. For Hal Finney's statement of the problem this section solves, see "Digital Cash & Privacy" (1993), https://nakamotoinstitute.org/library/digital-cash-and-privacy/, arguing that electronic payments create transaction dossiers unless spending is detached from account-level logs. Finney's "Protecting Privacy with Electronic Cash" (1993), https://nakamotoinstitute.org/library/protecting-privacy-with-electronic-cash/, gives the clearest period overview of the three-layer stack: public-key cryptography, anonymous messaging, and ecash.

[26]For Cashu, see https://cashu.space/.

[27]Cashu protocol specification: https://github.com/cashubtc/nuts (Cashu NUTs — Notation, Usage, and Terminology). Reference mint implementation: Nutshell, https://github.com/cashubtc/nutshell. Reference wallet: Cashu.me (web), https://cashu.me. The Cashu protocol was designed by Calle (https://github.com/callebtc) drawing directly on Chaum's 1982 blind-signature construction. For a technical walkthrough of the mint/token/redemption cycle and the operator-blindness property, see the Cashu documentation at https://docs.cashu.space/ and Calle's "Cashu: A Chaumian Ecash Wallet and Mint with Bitcoin Lightning" writeup, https://github.com/cashubtc/cashu/blob/main/docs/WHITEPAPER.md.

[28]For Fedimint, see https://fedimint.org/.

[29]Adam Back, Matt Corallo, Luke Dashjr, Mark Friedenbach, Gregory Maxwell, Andrew Miller, Andrew Poelstra, Jorge Timón, and Pieter Wuille, "Enabling Blockchain Innovations with Pegged Sidechains" (2014), https://blockstream.com/sidechains.pdf, is the foundational Liquid paper. Liquid Network documentation at https://liquid.net/, with the confidential-transactions construction documented at https://docs.liquid.net/docs/confidential-transactions-how-to and

in Andrew Poelstra, "Confidential Assets" (2017), https://blockstream.com/bitcoin17-final41.pdf. The Liquid Federation membership and governance are described at https://liquid.net/about/; Elements, the open-source codebase, is at https://github.com/ElementsProject/elements.

[30]Peter Todd, "Building Blocks of the State Machine Approach to Consensus" (2016), available at https://petertodd.org/2016/state-machine-consensus-building-blocks. The earlier proposal appears in Todd's November 2013 bitcoin-dev mailing list post on proof-of-publication consensus. For the RGB protocol's deployment of this approach, see the LNP/BP Standards Association's specifications at https://rgb.tech.

[31]Jonas Nick, Liam Eagen, and Robin Linus, "Shielded CSV: Private and Efficient Client-Side Validation," IACR Cryptology ePrint Archive, Paper 2025/068 (January 2025), available at https://eprint.iacr.org/2025/068. Prototype implementation at https://github.com/ShieldedCSV/ShieldedCSV.

[32]Monero encodes three privacy primitives at the base layer: ring signatures (Rivest-Shamir-Tauman 2001), stealth addresses, and Ring Confidential Transactions with Bulletproofs range proofs; the sixteen-member ring is the bounded anonymity set that decoy-selection attacks exploit and that Full-Chain Membership Proofs (FCMP++) are designed to replace with chain-wide anonymity. Monero project, https://www.getmonero.org/. Source at https://github.com/monero-project/monero. Monero Research Lab technical reports, https://github.com/monero-project/research-lab. Original ring signatures: Ronald L. Rivest, Adi Shamir, and Yael Tauman, "How to Leak a Secret," *ASIACRYPT 2001*. RingCT: Shen Noether, "Ring Confidential Transactions," IACR ePrint 2015/1098 (2015). Bulletproofs: Benedikt Bünz, Jonathan Bootle, Dan Boneh et al., "Bulletproofs: Short Proofs for Confidential Transactions and More," *IEEE Symposium on Security and Privacy* (2018). Bulletproofs+: Heewon Chung, Kyoohyung Han, Chanyang Ju et al., "Bulletproofs+: Shorter Proofs for Privacy-Enhanced Distributed Ledger," IACR ePrint 2020/735. Dandelion++: Giulia Fanti, Shaileshh Bojja Venkatakrishnan et al., "Dandelion++: Lightweight Cryptocurrency Networking with Formal Anonymity Guarantees," *ACM SIGMETRICS* (2018). On the empirical traceability of pre-RingCT Monero and the closed-set chain-reaction attack, Malte Möser, Kyle Soska, Ethan Heilman et al., "An Empirical Analysis of Traceability in the Monero Blockchain," *Proceedings on Privacy Enhancing Technologies* 2018, no. 3 (2018): 143–163, and Andrew Miller, Malte Möser, Kevin Lee, and Arvind Narayanan, "An Empirical Analysis of Linkability in the Monero Blockchain," IACR ePrint 2017/338, remain the standard references. On decoy-selection analyses and wallet-level responses, see Justin Ehrenhofer's survey work and the Monero Research Lab bulletins at https://www.getmonero.org/resources/research-lab/. On Full-Chain Membership Proofs, Luke Parker (kayabaNerve), "FCMP++: Full-Chain Membership Proofs for Monero" (2024),

https://www.getmonero.org/2024/04/27/fcmps.html, with specification and implementation at https://github.com/kayabaNerve/fcmp-plus-plus. On transaction-size and throughput comparison, see the Monero Research Lab "Empirical Results" bulletins and the broader comparative analysis in Harry Halpin and Ania Piotrowska, "Forensic Analysis of Privacy Coins" (2020), as well as ongoing per-block statistics at https://moneroblocks.info/ and https://mempool.space/ for the Bitcoin counterpart.

[33]On the Zcash transaction size and verification cost comparison, and the September 2018 Sprout spam attack: the attack is documented in Electric Coin Company retrospectives and Zcash community forums from September–October 2018; the bloated mempool and node slowdown are referenced in ECC's Sapling launch communications as part of the motivation for Sapling's proof-system redesign. On Sapling's efficiency improvements over Sprout, see Sean Bowe, Jack Grigg, Ian Miers et al., "Sapling," Electric Coin Company technical specification (2018), https://github.com/zcash/sapling-spec, and the Sapling protocol specification at https://zips.z.cash/protocol/sapling.pdf. On Orchard batch verification, see the Zcash protocol specification at https://zips.z.cash/protocol/protocol.pdf §5.4.

[34]On Ethereum's privacy architecture: Railgun protocol documentation at https://railgun.org/; Railgun smart contracts at https://github.com/Railgun-Community/. Aztec Network documentation at https://docs.aztec.network/; Aztec source at https://github.com/AztecProtocol/aztec-packages. Aztec launched its Alpha mainnet March 31, 2026; the launch announcement is at https://aztec.network/blog/announcing-the-alpha-network. A critical vulnerability in the Alpha v4 proving system was disclosed March 27, 2026, with a fix planned for a v5 release in July 2026; see https://aztec.network/blog/critical-vulnerability-in-alpha-v4. Aztec's predecessor product, Aztec Connect, stopped accepting deposits March 21, 2023, and shut down its sequencer March 31, 2024; all code is open-sourced. On the validity-vs-privacy distinction for ZK rollups, the ethereum.org documentation on ZK rollups states explicitly that rollups publish state data for every transaction processed off-chain to Ethereum, enabling independent reconstruction of rollup state; see https://ethereum.org/en/developers/docs/scaling/zk-rollups/. On Tornado Cash: U.S. Department of Treasury, OFAC designation of Tornado Cash, August 8, 2022, https://home.treasury.gov/news/press-releases/jy0916. Alexey Pertsev convicted in the Netherlands, May 2024. On Privacy Pools: Vitalik Buterin, Jacob Illum, Matthias Nadler, Fabian Schär, and Ameen Soleimani, "Blockchain Privacy and Regulatory Compliance: Towards a Practical Equilibrium," SSRN 4563364 (2023), https://ssrn.com/abstract=4563364.

[35]DarkFi is Amir Taaki's zero-knowledge smart-contract platform. Every contract is compiled to a zero-knowledge circuit, and transactions carry proofs

instead of readable input/output data, so arbitrary programmable logic executes under encryption. Project site and documentation at https://dark.fi/. Source at https://codeberg.org/darkrenaissance/darkfi. On Taaki's broader work in the cypherpunk tradition, see Carl Miller, "The Death of the Gods: The New Global Power Grab" (Windmill Books, 2018), ch. 8, and interviews collected at https://agorism.dev/.

[36]Zcash applies zk-SNARKs to shielded transactions across three proof-system generations (Sprout 2016, Sapling 2018, Orchard/Halo 2 2022); privacy-coin exchange delistings go back to Japan's 2018 FSA pressure, extended through Bittrex's January 2021 U.S. delisting, and accelerated in 2023–2024 as MiCA took shape; Monero and Zcash both lack the base-layer scripting that enables Bitcoin's Lightning and HTLC-based Layer 2 stack. Zcash project, https://z.cash/. Electric Coin Company, https://z.cash/. Zcash protocol specification at https://zips.z.cash/protocol/protocol.pdf. On the NU5 upgrade and Halo 2, see https://z.cash/blog/explaining-halo-2/. Halo 2 paper: Sean Bowe, Jack Grigg, and Daira Hopwood, "Recursive Proof Composition without a Trusted Setup," IACR ePrint 2019/1021. On the trusted-setup ceremonies, see Matthew Green, "Zcash Counterfeiting Vulnerability Successfully Remediated" (2019), https://z.cash/blog/zcash-counterfeiting-vulnerability-successfully-remediated/. On the scripting comparison, see Jameson Lopp's overview of Bitcoin Script primitives at https://www.lopp.net/bitcoin-information.html and the Monero and Zcash protocol specifications linked above; for the academic framing of the privacy-vs-programmability tradeoff, see Andrew Poelstra, "Confidential Assets" (2017), https://blockstream.com/bitcoin17-final41.pdf, which analyzes the structural conflict. On the history of privacy-coin exchange delistings: Japan's 2018 FSA pressure and the Coincheck response are covered in *Reuters* reporting from June–September 2018; ShapeShift's 2018 KYC transition and associated delistings at https://shapeshift.com/; Bittrex U.S. delisting announcement (January 2021) at https://bittrex.zendesk.com/; Kraken EEA Monero delisting (effective February 2024) at https://support.kraken.com/; Binance Monero delisting announcement (February 2024) at https://www.binance.com/en/support/announcement/; OKX privacy-coin delistings (late 2023–early 2024) tracked in *CoinDesk* and *The Block* coverage at https://www.coindesk.com/ and https://www.theblock.co/. On the Travel Rule's interaction with privacy coins, FATF Recommendation 16 guidance at https://www.fatf-gafi.org/.

[37]Mikhail Kudinov and Jonas Nick, "Hash-based Signature Schemes for Bitcoin," IACR Cryptology ePrint Archive, Paper 2025/2203 (December 2025), https://eprint.iacr.org/2025/2203. The paper argues that hash-based signature schemes are preferable for Bitcoin over lattice-based schemes (ML-DSA) because their security relies solely on hash-function assumptions already present in Bitcoin's design. It applies recent optimizations (SPHINCS+C, TL-WOTS-TW, and PORS+FP) to produce variants with substantially smaller signatures than

standardized SLH-DSA, and analyzes constraints specific to Bitcoin including key derivation, multi-signatures, and threshold signatures. Kudinov is at Blockstream Research; Nick is a Bitcoin protocol developer also at Blockstream Research and a co-author of the MuSig2 and BIP-327 work. For the SHRINCS scheme: C++ implementation at `https://github.com/BlockstreamResearch/shrincs-cpp`; Simplicity verifier at `https://github.com/BlockstreamResearch/shrincs-simplicity-verifier`; specification at `https://github.com/BlockstreamResearch/shrincs-specification`; parameter exploration scripts at `https://github.com/BlockstreamResearch/SPHINCS-Parameters`. The performance figures cited (1,092 bytes stateful at q=1, 4,396 bytes stateless) are from the SHRINCS-L Liquid performance report in the shrincs-cpp repository, benchmarked on an Intel Core i5 with 16 GB RAM; live Liquid transactions are linked in that report.

[38]Shor's algorithm breaks secp256k1 signatures on a sufficient quantum computer, placing Bitcoin funds at reused addresses and P2PK outputs at risk; BIP-360's Pay-to-Merkle-Root (P2MR) output type defends against long-exposure attacks by removing Taproot's key-path spend, with post-quantum signatures layered on top in a separate future proposal. BIP-360, version 0.11.0 (February 2026), "Pay-to-Merkle-Root," Hunter Beast, Ethan Heilman, and Isabel Foxen Duke, `https://github.com/bitcoin/bips/blob/master/bip-0360.mediawiki`. The BIP was previously named P2QRH (v0.1, September 2024) and P2TSH (v0.10, September 2025) before settling on P2MR; the changelog in the BIP itself documents this history. On the distinction between long-exposure and short-exposure attacks and the rationale for separating the P2MR output type from the post-quantum signature work, see the BIP-360 motivation and glossary sections. NIST post-quantum signature standards (all finalized August 13, 2024): FIPS 204 (ML-DSA, based on Dilithium), `https://csrc.nist.gov/pubs/fips/204/final`; signature sizes: ML-DSA-44 = 2,420 bytes, ML-DSA-65 = 3,293 bytes, ML-DSA-87 = 4,595 bytes (FIPS 204 Table 1). FIPS 205 (SLH-DSA, based on SPHINCS+), `https://csrc.nist.gov/pubs/fips/205/final`; signature sizes range from 7,856 bytes (SLH-DSA-SHA2-128s, small/slow) to 49,856 bytes (SLH-DSA-SHA2-256f, large/fast) across the twelve parameter sets (FIPS 205 Table 2). FALCON (the compact lattice-based scheme, Falcon-512 approximately 666 bytes, Falcon-1024 approximately 1,280 bytes) was selected by NIST for ongoing standardization; as of the time of writing, no FIPS number had been assigned and the standard had not been published. NIST post-quantum cryptography project overview: `https://csrc.nist.gov/projects/post-quantum-cryptography`. On the current consensus timeline for a cryptographically relevant quantum computer capable of breaking secp256k1: the dominant estimate in the research community and among security agencies is 10–20 years, with the NSA's CNSA 2.0

suite mandating transitions by 2030–2035; a 2024 paper by Alice & Bob (https://alice-bob.com/blog/computing-256-bit-elliptic-curve-logarithm-in-9-hours-with-126133-cat-qubits/) substantially reduced prior resource estimates, calculating that 126,133 cat qubits could in principle break ECC-256 in 9 hours, though no machine near this qubit count exists. For ongoing bitcoin-dev and DelvingBitcoin discussion of quantum migration, see https://gnusha.org/pi/bitcoindev/ and https://delvingbitcoin.org/. On the fraction of Bitcoin supply at quantum risk, see Deloitte, "Quantum Computers and the Bitcoin Blockchain" (2021). Bitcoin Optech post-quantum coverage at https://bitcoinops.org/.

21

DECENTRALIZED SOCIAL INFRASTRUCTURE

> "The simplest open protocol that is able to create a censorship-resistant global 'social' network once and for all."
>
> fiatjaf, Nostr Protocol README (2020)[1]

Social coordination has concentrated into a handful of platforms that own the identity layer. This is the core vulnerability: when the platform holds the account, a ban does not just restrict speech, it destroys the social graph built around it. Followers, reputation, post history, and audience relationships are erased together in a single administrative action.

Nostr is a protocol that moves identity control to the user. An account is a keypair the user generates and controls; it is not an allocation granted by any operator. Relays store and forward content but hold no authority over the identity. The protocol addresses portability and censorship resistance first; privacy properties follow from the architecture and are treated in the sections below.

21.1 The Problem with Centralized Platforms

What makes social platforms distinctively dangerous is identity capture. Two corporate entities (Meta and Alphabet) control the primary social graph infrastructure for the majority of U.S. internet users. As of 2025, YouTube reaches 84% of U.S. adults and Facebook

71%; Meta alone owns four platforms each with more than one billion monthly active users.[2] A coordinated decision across two or three corporate actors can effectively terminate a user's public presence across the entire accessible internet. Chapter 11's analysis of corporate surveillance describes the accompanying extraction model: behavioral profiling and social graph mapping, tied to state entanglement through data sharing.

The Twitter Files (December 2022–March 2023) documented the state-entanglement mechanism in operational detail: the FBI maintained a formalized channel for submitting account review requests to Twitter's Trust and Safety team, and U.S. Central Command ran covert influence operation accounts that Twitter's Site Integrity team had whitelisted.[3] Formal orders were unnecessary; routine communication between government agencies and platform trust-and-safety teams sufficed.

De-platforming Risk

De-platforming is complete removal from the platform and everything it holds on the user's behalf. In August 2018, Apple, Facebook, YouTube, and Spotify removed Alex Jones and InfoWars in a coordinated 12-hour window; Jones lost a YouTube channel with approximately 2.4 million subscribers and over 1.5 billion archived views simultaneously with his removal from every other major platform.[4] In January 2021, Twitter removed over 70,000 accounts following the Capitol breach, and Amazon Web Services terminated Parler's hosting, taking the entire platform offline and wiping account data for its roughly 15 million registered users in a single infrastructure action.[5]

No advance notice preceded any of these terminations. Emergency safety provisions allow immediate permanent action with no grace period for content export or contact backup. The audience relationship a creator spent years building is deleted alongside the account.

The Portability Problem

Creators never possess their own audience. When a user follows an account on a centralized platform, they provide their contact information to the platform. The creator cannot export that list. GDPR Article 20 grants users the right to download their own personal data in a machine-readable format, and all major platforms now comply; follower contact data belongs to the followers, not to the creator.[6] You can download your own posts; you cannot download the contact information of your two million followers.

An email newsletter operator with 100,000 subscribers can migrate from one provider to another by exporting a CSV. No equivalent mechanism exists for social platform followers. The social graph is the platform's proprietary asset.

Impersonation and Content Integrity

Centralized platforms provide no cryptographic guarantee that content attributed to an account was produced by its owner. The platform authenticates login sessions, not statements. An administrator with database access can edit or delete a post without any visible record of the change. A screenshot of a tweet is not evidence that the tweet was ever sent; the platform can silently alter or remove content and users have no mechanism to detect it.

Impersonation compounds this. A verified badge signals that a platform has confirmed some relationship between an account and a real-world identity, but the verification is the platform's assertion, not a proof the user can independently check. Platforms have transferred verified badges between accounts and sold them outright. When the authentication layer is controlled by the platform, users have no way to distinguish authentic statements from fabricated ones except by trusting the platform's own records.

Lock-in Effects

The network effects and lock-in dynamics analyzed in Chapter 11 explain why users remain on platforms they dislike. Individual

switching does not bring the network along, and the mechanisms that enforce this go deeper than simple network size.

Content archives function as sunk cost. Everything a user has posted exists only inside the platform's database. A user who leaves forfeits that record entirely; there is no export path that preserves the social layer around the content. The posts exist as files, but the reply chains and engagement history that give those posts meaning do not survive export.

Algorithmic amplification compounds the effect. Platforms weight reach by engagement history on that specific platform. A creator who built an audience of one hundred thousand over five years cannot transfer the engagement signals that drove that distribution. Starting over on a new platform means starting at zero audience reach regardless of content quality, because the new platform has no prior engagement data to feed into its recommendation system.

API restrictions prevent third-party migration tooling from bridging the gap. Platforms can and do revoke or rate-limit API access that would allow automated follower contact or content migration. Twitter's 2023 API pricing change, which raised costs by several orders of magnitude, effectively shut down independent research tools and third-party migration clients that had been running for years.[7] What remains technically possible is routinely made economically infeasible.

The switching cost is collective because audience relationships require bilateral action. A creator can announce a move to a new platform; whether the audience follows is a separate coordination problem that the creator cannot solve unilaterally. The platform's hold on the user is enforced not by contract but by the coordination failure that would need to be overcome to escape it.

Algorithmic Opacity

What users see is determined by proprietary ranking algorithms: closed-source, unaudited, set by the platform's commercial interests. The feed is not a chronological record of content from accounts a user chose to follow; it is a filtered and reordered selection produced by code whose optimization targets serve the platform, not the user.

Shadowbanning is the covert variant of this control. A shadowbanned account remains nominally active (the user can still post and appears visible), but the platform's distribution systems suppress the content so that replies go unanswered and follower timelines never surface the posts. The user receives no notification and has no mechanism to appeal. Twitter's own internal documents, released through the Twitter Files, confirmed that the platform maintained separate internal visibility tiers that differed from what users saw on their own profiles.

The opacity is structural. Platforms have no incentive to publish ranking criteria because the criteria are also advertising targeting levers. The same signals that determine who sees a post determine which ads appear alongside it. Opening the algorithm to inspection would expose the commercial logic of the feed. Users interact with a system whose rules they cannot read and can only infer from observed behavior, and even that inference is unreliable because the algorithm changes continuously without notice.

21.2 Nostr: The Protocol Solution

Simple Protocol, Complex Network

Nostr (Notes and Other Stuff Transmitted by Relays) is a protocol, not a platform.[8] It specifies how messages are formatted and how clients and relays communicate. Anyone can implement clients and operate relays.

The core protocol is remarkably simple. All content takes the form of "events": JSON text files with a standard format containing content, timestamp, public key of author, signature, and tags for metadata. An event is nothing more than a signed text file. This radical simplicity means events can be stored and transmitted through any channel, then processed by any software that can read JSON and verify signatures. Relays are servers that store and forward events, though they may accept or reject even validly signed events according to local policy and some require authentication extensions such as NIP-42. Clients are applications that users interact with; they

fetch events from relays and display them, then create new events that users sign.

This simplicity enables permissionless innovation. A competent developer can build a client in days. A builder needs no API keys or approval process, and no terms of service govern access beyond the protocol itself. The result is thousands of clients with different focuses: mobile, desktop, long-form content, images, video, chat, marketplaces.[9] No one controls who can build clients or operate relays.

Notes and Relays Architecture

The architecture separates user control from client and relay roles. Users hold private keys and create signed events; clients provide the interface and manage relay connections; and relays store and distribute events without being authoritative over any of them. Users can connect to multiple relay simultaneously, and content is replicated across the relays they choose.

No single relay is authoritative. One relay can refuse a user or fail outright, but others can still carry the content. Users choose relays based on policy and community fit.

The NIP System

Nostr Implementation Possibilities (NIPs) are the mechanism by which the protocol grows.[10] A NIP is a written specification: a document that defines a new event kind, a new field, a new behavior, or a new interaction pattern, written precisely enough that any developer can implement it independently and interoperate with any other implementation. The specification lives in a public repository. Anyone can write one.

The process is open by design. A developer who wants to add functionality to Nostr writes a NIP describing exactly how it works (the event kind number, the required fields, the expected client and relay behavior) and submits it. Other developers read it, argue about it, propose amendments, and decide whether to implement it in their own clients and relays. There is no committee

that approves or rejects NIPs for inclusion in the protocol. Adoption is the only vote that counts. A NIP that solves a real problem gets implemented; one that does not gets ignored. The repository accumulates both, and the distribution of implementations reflects developer judgment, not the preferences of any central body.

NIP-01 defines the base protocol: the event format, the relay communication model, the subscription syntax. Everything built on top of it is optional. Human-readable identifiers mapped to public keys via DNS come from NIP-05, giving users addresses like `alice@example.com` that resolve to the cryptographic key of Alice. Lightning zaps, the mechanism by which users attach Bitcoin micropayments to events, are defined by NIP-57. Relay authentication, allowing relays to require clients to prove key ownership before accepting content, comes from NIP-42. Encrypted direct messages are NIP-44. Each of these started as a document someone wrote and published, inviting independent implementation.

The consequence is that no single entity controls what Nostr can do. An app that wants to add a feature ships a NIP and invites adoption. A security researcher who finds a flaw in an existing construction writes a replacement NIP and makes the case for migration. The protocol surface expands or contracts based on what developers build and what users use, with no permission required from anyone.

21.3 Keys as Identity: Cryptographic Sovereignty

Public Key as Identity

In Nostr, your identity is your public key. Generating a keypair locally is the entire registration process; no server grants it and no server can revoke it. No username or phone number enters the loop.

The public key is your permanent identifier. Every event you publish is signed with your private key, and that signature cryptographically ties the event to your public key. Anyone receiving the event can verify the signature without contacting any authority. Identity operates by possession: whoever holds the private key controls the identity, and no third party can override that.

The contrast with account-based systems is architectural. On a traditional platform, your account is a row in a database the platform controls, editable or deletable at administrator discretion. On Nostr, your identity is a mathematical object that exists independently of any database. A relay can refuse to carry your events, but it cannot modify them or invalidate your key. There is no authority with the power to suspend the identity itself.

User-Controlled Identity and Profile

Once a user creates a keypair, they can self-declare their entire identity through a kind 0 event. This special event type contains a JSON object with profile metadata: username, display name, biography, profile picture, banner image, website, Lightning address for receiving payments, and any other fields the user wishes to include. The user signs this information with their private key and publishes it to relays.

No third party can stop a user from choosing their desired identity. No approval process or content review governs what username or profile picture one may select. If one relay rejects the event, others may still accept it. The identity exists the moment the user signs it.

This inverts the traditional relationship between users and platforms. On centralized services, the platform owns the namespace and grants users permission to occupy a slot within it. On Nostr, users create identities that exist independently of any infrastructure. The profile is just a signed text file that any software can read and display.

Key Management Challenges

Key control creates responsibility and introduces challenges absent from traditional account-based systems.

Key loss is permanent. Lose your private key, lose your identity. There is no "forgot password" recovery or customer support path, and no secondary verification method. The accumulated reputation and history associated with that public key, become inaccessible.

Key compromise is equally permanent but in a different direction. If an attacker obtains your private key, they can impersonate you indefinitely. Account-based systems let administrators lock a compromised account and reset credentials through alternative verification; Nostr has no such recovery path. The attacker with your key is cryptographically indistinguishable from you: they can post and sign as you with no authority to appeal to. Compromise is permanent identity theft.

Key rotation presents its own difficulties. Traditional systems allow password changes that maintain account continuity. As of now, Nostr has no widely adopted mechanism for rotating to a new key while preserving identity continuity. Moving to a new key means starting over: new public key, zero followers, no history. Some clients support key rotation announcements where the old key signs a message endorsing the new one, but adoption is inconsistent and many clients do not recognize these transitions. Users who suspect compromise face a choice between continuing with a compromised key or abandoning their accumulated social capital.

Users face tradeoffs in managing these risks. Self-custody offers maximum control along with maximum responsibility, while custodial solutions allow services to hold keys and enable recovery at the cost of trust in the custodian. Threshold schemes split keys among multiple parties so that recovery requires a subset of them to cooperate. These tradeoffs are inherent to self-sovereign systems. Nostr makes them explicit instead of hiding them behind platform-controlled "accounts."

Social approaches can soften the problem without solving it. A user who still controls a compromised key can publish a plain warning that the key should no longer be trusted, while friends and counterparties can attest that a new key is the rightful successor. That helps preserve continuity and warn the wider network, but it depends on client support, visible conventions, and social verification. Identity recovery on Nostr remains a live problem.

21.4 Relay Architecture and Market Dynamics

Anyone Can Operate a Relay

Running a Nostr relay requires only a server and the relay software. No permission or approval is needed; the relay software is open source and the protocol is public.

Individuals run personal relays; communities run specialized relays; businesses run commercial relays. The barrier is low enough for hobbyists and individual developers, not only well-funded operators. A relay runs on a phone or an old laptop as readily as on a $3-per-month VPS; storage requirements are modest because each relay stores only the events it chooses to accept, not the full network history.

Relay Competition on Service Quality

Relays compete across several dimensions. Reliability covers uptime and response latency; a relay that drops connections or loses events drives users to alternatives. Storage policy determines how much history a relay retains and which event kinds it indexes; relays that prune aggressively lose users whose older content matters to them. Feature sets vary: some relays offer full-text search, others provide analytics dashboards or specialized handling for particular event kinds like long-form articles or marketplace listings. Policy determines what content the relay accepts or rejects, which attracts communities that share those values. A relay that enforces strict content standards pulls in users who want that environment; a permissive relay attracts others.

Users respond to these differences by configuring multiple relays and weighting them differently. A user might publish to a high-uptime paid relay for broad reach while also publishing to a community relay that serves their specific audience. The client manages the distribution; the user sets the policy. Relays that serve users well retain them; those that do not are dropped from the list.

Paid vs. Free Relay Models

Relay funding takes several forms. Operator-subsidized relays offer free access but bear the cost themselves; paid relays charge users directly, which both funds the operation and raises the cost of spam; freemium relays give free access for basic use and charge for premium features. Paid relays solve the sustainability problem while creating incentives for quality service. Free relays serve users who cannot or will not pay but face sustainability challenges at scale.

The payment mechanics vary by relay. Subscription relays charge a flat fee, commonly denominated in sats per month, topped up via Lightning payment; the relay checks the account balance before accepting events and stops serving the account when it runs dry. Per-event relays take a different approach: each uploaded or downloaded event costs a small ecash token, making the pricing granular and the cost directly proportional to usage. Ecash tokens can be minted and redeemed without revealing payment identity, which preserves some user privacy at the relay layer. Relay operators who prefer conventional billing can accept fiat payments through standard payment processors; the relay software typically abstracts over the payment method, treating a valid subscription credential as the access condition regardless of how it was funded. The spread of payment options reflects the absence of any central relay authority: operators choose their own business model, and users choose relays whose terms suit them.

Current Centralization Tendencies

Despite decentralized design, Nostr exhibits centralization tendencies in practice. A few large relays carry most traffic because users default to well-known ones, and content discovery routes through a handful of indexing services. A few clients attract most users because network effects favor familiar software. The pattern reflects ordinary market dynamics: network effects drive coordination onto common infrastructure regardless of whether the underlying protocol enforces it. What the protocol does change is the consequence of that coordination. On centralized platforms the

coordination point becomes the platform, and the platform becomes the only place users can exist. On Nostr the coordination point can shift as users move between relays and clients without losing their identity, so exit remains possible even when any single relay or client dominates for a time.

The question is whether market competition will maintain sufficient alternatives to prevent reconcentration. The protocol enables decentralization; market outcomes determine whether decentralization persists.

21.5 Reputation Without Central Authority

Web of Trust Models

Without central verification, how do users evaluate credibility? In web of trust models, users vouch for others and trust propagates through the network; if you trust Alice and Alice trusts Bob, you have some reason to consider Bob credible. Follow graphs reveal community structure, and users followed by people you trust are more likely trustworthy. Explicit endorsements create additional reputation signals. The underlying idea is old: reputation grows through the communities a person operates in. What changes is that cryptographic signing makes the attestations explicit and verifiable by anyone who receives them.[11]

WoT is already deployed as a spam filter at the relay layer. The wot-relay implementation filters events by whether the author falls within a configurable hop distance of the relay operator's follow graph; events from accounts outside the neighborhood are dropped without the relay needing to evaluate content.[12] The Zapstore app store uses the same signals for software discovery: apps published by accounts within your trust graph surface preferentially over apps from strangers.[13] These are practical deployments of WoT as an anti-spam and curation mechanism, operating without any central registry.

Follows and Interactions as Reputation Signals

Reputation emerges from observable behavior over time. A follower count suggests perceived value at a given moment, while replies, reposts, and other reactions signal how readers evaluate specific pieces of content. Established accounts generally carry more weight than fresh ones because consistent behavior across a long record is expensive to manufacture, and the longer a key has operated under one persona the harder it becomes to discard and replace that persona without losing everything accumulated under it. None of these signals is infallible, and all of them are partially gameable, but together they provide credibility information without any central authority deciding in advance what counts as credible.

Verification Without Centralized Checkmarks

Nostr's NIP-05 enables verification through DNS.[14] Users can link their Nostr identity to a domain they control: "user@example.com" where example.com confirms the association.

Verification operates by domain control, not platform decision. Organizations verify employees by hosting their identities, while individuals verify themselves using personal domains. No central authority decides who is "verified."

The trust is in the domain, not in a platform checkmark. This distributes verification authority instead of concentrating it.

Verification Across Anonymous Channels

The trust graph becomes more useful when it crosses channels. Suppose someone messages you on an anonymous channel and claims to be a developer or merchant you know from Nostr. The immediate problem is not abstract reputation but continuity: does the person on this channel control the key whose public history you already know?

Challenge-response can answer that question. One side generates a nonce, then sends the challenge through the anonymous channel. The other side signs the challenge with the known Nostr key and returns the signed response, ideally through a one-time

encrypted relay message or by direct copy-paste if relay transport is unavailable. The verifier checks the signature, then checks the social graph around that key. Who follows it? Does it carry a NIP-05 identifier? Does its posting history match the claimed identity?

This proves continuity between the anonymous channel and the reputational identity. It does not prove innocence or commercial reliability. Those are larger institutional questions. But it does solve one real problem: how to verify a known pseudonym across channels without demanding legal identity.

The continuity question has a bounded answer. Protocol-level reputation signals and protocol-level proofs can carry trust only so far, and the wider questions of counterparties and dispute belong to the institutional layer taken up in Chapter 24.

Trusted Assertions: Structured Reputation Signals

NIP-85 formalizes reputation as a signed, structured data product.[15] A trusted assertion is a parameterized replaceable event published by a declared service provider, covering a subject identified by public key, event ID, addressable coordinate, or NIP-73 external reference. Each assertion carries computed values as structured tags: a rank score from 0 to 100, follower count, total zap amounts received, report counts, and inferred topic clusters derived from the subject's posting history. The provider signs the assertion with a key specific to the algorithm that produced it, so the signer is the algorithm and its declared methodology, not the individual who operates the provider key.

Users opt in to specific providers by publishing a kind 10040 event listing the provider keys they trust, optionally encrypted so the choice of provider is not itself public. Clients then retrieve assertions from the declared providers and use them to rank or filter content without performing the heavy graph computation locally. The result is a clean separation between reputation production and reputation consumption: providers compete on methodology and accuracy, clients consume signed outputs without running full WoT traversals, and users control which methodologies govern what they see. Nothing in the design requires any provider to be authoritative

for all users, and no central registry decides whose assertions count. The same key infrastructure that authenticates posts authenticates reputation claims, and the same user-controlled trust lists that govern relay selection govern whose reputation signals apply.

21.6 Moderation as Market Service

No Protocol-Level Moderation

Nostr has no protocol-level content moderation. The protocol transmits signed events and makes no judgment about their acceptability; that evaluation is left entirely to the layers above it. The absence is deliberate: protocol-level moderation would require some authority to define what is acceptable and enforce those definitions across the network, which would centralize exactly the power the protocol was designed to prevent. What remains instead is a division of labor across relays, clients, and user-selected curation services, each operating independently and competing for users on the merits of their policies.

Relays filter through acceptance policies and key-level blocking. Nothing in the protocol compels any relay to carry any event; each operator decides what they are willing to host. The same event can be accepted at one relay while rejected at another. A user whose content is refused at one relay publishes to another and loses only reach at that specific host, not access to the network.

Clients filter what they display through muting and algorithmic ranking, among other controls. Because clients run on the user's own device, the filtering is a user choice, not a platform decree. Users who dislike a given client's filtering approach can switch clients without losing identity or social graph.

Between the relay and the client sits a market for curation. Blocklist providers offer curated lists of accounts to filter, spam filters remove unwanted content, community moderators maintain local standards, and algorithmic feed services package ranking as a product. Users choose which curation services to subscribe to, and providers compete on quality and values. The result is moderation that functions without a single authority deciding what everyone

sees: each user composes their own experience from the relays and curation services that suit them, and disagreement about moderation is resolved by letting different users make different choices.

21.7 Beyond Social Media: One Protocol, Many Payloads

The same signed-event-plus-relay architecture that carries short social posts can carry anything else a developer wants to define, and the NIP extension system has been used to do exactly that. A long-form article, a classified listing, a live video stream, a wiki entry, an audio room, an encrypted group message: each is a Nostr event of a specific kind, routed through the same relays and authenticated by the same public keys. The content surface is wide; the mechanics underneath do not change.

That continuity is the interesting part. A writer who posts short notes under one keypair can publish long-form essays under the same keypair, and readers follow both through the same social graph.[16] A developer who has built social reputation can ship signed software releases that users discover through the same web of trust that every other Nostr application runs on, closing the app-store gatekeeper out of the loop entirely. A merchant whose public history is visible on the network can list goods in a peer-to-peer marketplace where the identity anchor is the public key and not a platform account, with Lightning handling payments and private messaging handling the rest.[17] A streamer can broadcast video whose discovery and tipping ride on the same relays that serve their ordinary posts, with social interaction woven into the same event stream.[18] Multiple authors can publish competing wiki entries on the same topic, and readers choose whose version to trust using reputation signals they already use for everything else.[19] Real-time spaces like HiveTalk and Nostr Nests authenticate participation through the user's existing keypair and integrate Lightning for access control or tips.[20] The catalogue is long and getting longer, but the shape of each case is the same.

Reputation is what ties this together. What makes the range possible is that a single identity and relay network serve all applications at once, with one social graph underneath them all, rather than each application duplicating the infrastructure independently. A keypair that has accumulated trust in one context carries that trust into the next, and the user can separate identities by context when separation is what they want. The protocol does not know what any payload means; it only knows the payload is signed and that some relays are willing to carry it. Everything else is a matter of which clients implement which event kinds and whether users find those clients useful enough to attract other developers adding the feature to their own clients.

Private-messaging schemes extend the same infrastructure. Encrypted direct messages (NIP-17 using NIP-44 encryption) and the Marmot protocol for scalable encrypted group messaging run as specialized event kinds on the same signed-event-plus-relay architecture. Public permissionless uses and private encrypted uses are both specializations of the same system, and the range visible today is a small fraction of what the protocol admits. That breadth comes with privacy properties the architecture does not provide by default, which the next section addresses directly.

21.8 Privacy Limitations

Pseudonymous, Not Anonymous

Nostr provides pseudonymity, not anonymity. Your public key is a persistent identifier. Every event you sign links to that key. Over time, behavioral patterns, writing style, timing, and social graph connections accumulate into a profile that may be deanonymizable.

Creating fresh keys provides unlinkability to previous identity but sacrifices accumulated reputation. A user who abandons a key to escape surveillance loses the follow graph, posting history, and web-of-trust standing that took time to build. The tradeoff is not a design flaw but a structural property of any persistent identity system: the same continuity that makes reputation legible also makes behavior traceable.

Social Graph Exposure

Follow lists, reposts, replies, and zaps all reveal social connections in the clear. Even an observer who reads no content at all can map community structure by watching who interacts with whom, at what frequency, and through which relays. The follow graph is published as a kind 3 event and replicated across every relay the user publishes to; anyone who queries those relays gets the full list. Reaction patterns add resolution: a user who consistently zaps and replies to a small cluster of accounts reveals affiliation even if every message in that cluster is encrypted.

Some clients support encrypted follow lists under NIP-51, but adoption is uneven and most users publish their social graph in the clear without recognizing what that exposes.[21] The exposure is not incidental to Nostr's design; it is the mechanism by which web-of-trust filtering and content discovery work. Users who need graph privacy must configure it deliberately and accept that doing so degrades the discovery and reputation features that depend on the public graph.

Relay Operators See Everything

Relays receive events in cleartext and have full visibility into the traffic they handle. A relay operator can see which public keys post which content, which IP addresses request which events, the timing and frequency of all interactions, and the emerging shape of the social graph among their users. The visibility follows from using relays as the storage and delivery layer without a separate anonymization layer underneath.

Legacy NIP-04 direct messages used AES-256-CBC with no message authentication code, making ciphertext malleable, and derived the encryption key from a non-standard ECDH construction that diverged from the libsecp256k1 default. NIP-44 replaced it with ChaCha20 plus HMAC-SHA256 authenticated encryption and proper HKDF key derivation; a Cure53 audit of NIP-44 published in December 2023 confirmed the design.[22] NIP-44 is an encryption primitive, not a complete messaging protocol. Used alone, it would

leave sender identity visible to relays. NIP-17 composes NIP-44 encryption with NIP-59 gift wrapping to address that gap: the outer event (kind 1059) is signed by a fresh ephemeral keypair, its `created_at` timestamp is randomized up to two days in the past, and the `p` tag identifies only the recipient.[23] A relay serving NIP-17 messages sees neither the sender's identity nor the real timestamp, only when the event was first delivered to the relay. Content and sender are both hidden from the relay.

The remaining exposure is network-level. The relay still sees the IP address that connected to deliver or retrieve the gift wrap, and users who publish to a dedicated DM inbox relay (kind 10050) reveal that relay as the place to watch. Someone with observation points between the user and that relay can correlate connection timing with message delivery even without reading the content.

Public Reach as the Default Posture

The standard Nostr workflow broadcasts notes to several relays so they reach as wide an audience as possible, and any relay that carries an event will serve it to anyone who queries for it. A user who posts a note expecting it to circulate only among their followers has in fact handed it to every relay on their publish list. The encrypted alternatives described above (NIP-17 direct messages, Marmot group messages) exist as protocol extensions but require deliberate configuration and client support. Users who arrive expecting private-by-default behavior will expose more than they intend unless they pick those alternatives explicitly.

Marmot: MLS over Nostr

Chapter 17 introduced Messaging Layer Security, the IETF standard for scalable group encryption. MLS on its own provides forward secrecy, post-compromise security, and O(log N) group operations, but it deliberately leaves two jobs to the surrounding system: authenticating who a group member is, and delivering ciphertext without leaking the social graph. The Marmot protocol uses Nostr to do both.[24]

Marmot treats the Nostr keypair as the MLS Authentication Service: a user's long-term Nostr identity key anchors their MLS credential, so there is no phone number, email address, or certificate authority in the loop. It treats Nostr relays as the MLS Delivery Service: any relay that accepts the protocol's event kinds can carry the traffic, and no specific relay is required. Three event kinds cover the protocol. KeyPackage events (kind 30443) are published in advance so others can add a user to a group asynchronously, mirroring Signal's prekey model. Welcome events (kind 444) are delivered to newly added members wrapped in NIP-59 gift-wrapping, which uses onion-style layering to hide the sender from the relays that route them. Group messages (kind 445) carry the actual MLS payload.[25]

The reference implementation is the Marmot Development Kit (MDK), a Rust library built on OpenMLS; the flagship client White Noise ships on iOS and Android and reuses the same Rust backend through a terminal client for desktop use. Least Authority published three reviews of the work between November 2025 and April 2026, covering the specification, the MDK library, and the whitenoise-rs backend.[26]

Two design choices give Marmot metadata protection that MLS alone does not provide. First, the MLS signing key inside KeyPackages and leaf nodes is distinct from the user's long-term Nostr identity key, so a later compromise of the Nostr key does not expose past or future group traffic, not even which groups the user was a member of. Second, every group message is published under a fresh ephemeral Nostr keypair instead of the user's identity key, and the group's public-facing identifier is a rotatable pseudo-ID instead of the real MLS group ID. A relay operator watching the event stream sees ciphertext tagged with a rotating handle, with no persistent sender or stable group label to cluster on. Combined with MLS's native forward secrecy and post-compromise security, this closes most of the gap between content encryption and metadata privacy.

The residual leak is in the subscription itself. A client that participates in several Marmot groups subscribes to each group's rotating `h` tag in one REQ filter so a single connection can deliver

traffic from all of them. The relay reading that filter sees the set of group identifiers the client follows, and a relay that serves many users can cluster sessions by the overlap in their followed groups even when each group identifier rotates and each event is signed by a fresh ephemeral key. Two users whose subscription sets overlap on the same handful of groups are visibly members of the same groups, and a user who subscribes to a distinctive cluster of groups identifies themselves across sessions even when their transport runs over Tor. The structural fix is private information retrieval: with PIR over Nostr relays, a client could pull events for a set of groups without revealing which set, severing the only remaining traffic-analysis handle Marmot leaves to a relay.

Take a standards-track group encryption protocol, compose it with a transport that no single entity can shut down, and the result is scalable, forward-secret, metadata-hiding group messaging whose identity layer is self-sovereign and whose delivery layer is no one's property. It is what the progression described in Chapter 17 has been pointing at, minus the PIR layer that would close the subscription leak.

Network-Level Metadata

Every WebSocket connection to a relay carries the client's IP address, and relay operators may log it alongside the public key being used. Without Tor or a VPN, relay operators and any network observer positioned between the client and relay can link a public key to a network location and, from there, to a physical address or ISP account. The problem compounds across relays: a user who connects to five relays from the same IP has correlated five independent operator logs into a single identity profile, even if the user otherwise takes care to separate their activity across those relays.

The protocol imposes no anonymization requirement and provides no built-in transport privacy. Users who want IP-level protection must route connections through Tor or a trusted VPN before they reach the relay, and they must do so consistently; a single unprotected connection to a relay that has seen the same public key

over Tor is sufficient to break the pseudonym. The responsibility sits entirely with the user.

What Clients Can Do: Tor, ZK Authentication, Private Retrieval

Privacy on Nostr is a client problem, since the protocol leaves it to the application layer. The simplest mitigation is to route every relay connection through Tor or a trusted VPN, so the IP address the relay logs is a Tor exit or a VPN egress instead of the user's home address. A client that does this consistently from the first connection prevents the operator from ever associating a public key with a network location. Several Nostr clients ship Tor support either as a built-in transport or by deferring to a system-wide Orbot or Tails configuration.[27] The cost is latency and the discipline of using the protected path every time, since one unprotected connection links the pseudonym back to its real network home.

Deeper protocol extensions are possible but not yet deployed at scale. Several building blocks from the cryptographic literature address what a relay currently sees about a connected client: the authorized user's identity on the line and the content of their subscription.[28]

Zero-knowledge group membership proofs hide which authorized user is connecting. Semaphore, a construction battle-tested at scale in Worldcoin, lets a relay publish a Merkle root of authorized public keys and lets a client prove "I am one of these members" without revealing which one. A per-epoch nullifier derived from the user's secret key provides spam protection: two proofs from the same key in the same epoch produce the same nullifier, so the relay can reject duplicates without learning the underlying identity. Mobile proofs take a few seconds to generate and milliseconds to verify, and the libraries exist today. The missing piece is a NIP and adoption.

Private information retrieval addresses the more difficult problem of hiding what a client is fetching. In a PIR scheme, the client encodes a query as a vector and the relay performs a computation over its full event database that returns the requested record without learning which index was queried. Schemes like SimplePIR and

FrodoPIR have brought single-server PIR into the range of practical deployment for bounded query patterns, with sub-second latency on modestly sized stores. Server computation scales with database size, and communication overhead runs an order of magnitude above the size of the data retrieved. That cost makes PIR viable for high-value queries (fetching a specific encrypted inbox, retrieving a particular thread) rather than for firehose subscriptions that pull thousands of events. Adapting PIR to Nostr's subscription-style query model is an open research problem, not a deployment-ready solution.

These mitigations stack. Tor hides the IP. ZK authentication hides which authorized user is on the line. PIR hides what they are reading. None is a complete solution alone, and the deployable layer today is Tor at the transport level, with ZK authentication reachable through a new NIP. PIR remains the harder open problem, and a privacy-conscious client today should at minimum use Tor by default and treat the relay-visible query surface as still exposed.

The Tradeoff

Nostr optimizes for censorship resistance and user control, not privacy. Users own their identity and cannot be deplatformed, but their activity is broadly visible. This is a reasonable tradeoff for public social communication, where the goal is often reach, not concealment. Users requiring strong privacy should use Nostr cautiously and supplement it with privacy tools, or use purpose-built private communication systems for sensitive interactions.

21.9 Why Alternatives Fall Short

Mastodon: Federated but Server-Dependent

Mastodon uses federation: independent servers communicate through the ActivityPub protocol.[29] Users register on servers (instances), which relay content between each other.

Identity is server-dependent. Your identity is "user@server.example" (the server is part of the address). If the server shuts down or

bans you, that identity is gone. Mastodon added account migration in 2019, which allows users to move to a new server and carry their follower count, but content does not migrate: posts and reply threads remain on the original server and are inaccessible if that server goes offline. Migration also requires the old server to still be running and cooperating; a server that shuts down abruptly or bans you first makes migration impossible.

The practical consequences are visible. When large Mastodon instances have shut down (mastodon.social temporarily closed registrations in 2022 under load, and dozens of smaller instances have shut down with little notice), their users lost content archives and often follower relationships. The federated model distributes operator control across many servers, which is better than a single platform, but the identity is still a credential the server operator issues. Users who migrate frequently for better moderation or reliability lose their content history each time. Mastodon distributes control among server operators; it does not give users control of their own identity.

Bluesky: Credible Exit, Not Current Decentralization

Bluesky uses an open protocol (AT Protocol) with a different architecture than Nostr.[30] Where Nostr events are simple signed text files that any relay can store independently, ATProto uses a "shared heap" model: relays must aggregate and index the entire network's data to function. This architectural choice creates inherent centralization pressure regardless of protocol openness.

The resource requirements reflect this difference. A full ATProto relay requires multi-terabyte storage that grows at roughly eighteen gigabytes daily, and a full App View, which is the service that renders feeds, adds hundreds of thousands of dollars of indexing infrastructure on top. For most of the network's history, Bluesky PBC operated the only full-network relay; Bluesky still runs the dominant firehose and in fact operates multiple full-network relay instances of its own across data centers. In late 2025 that changed when Blacksky brought up a second independent full-network relay on its own Rust implementation (rsky-relay), confirmed by Bluesky PBC's own protocol check-in in October 2025.[31] Thousands of per-

sonal data servers have launched alongside these relays, but most host only one or two accounts and still depend on the larger relays and App Views to reach the broader network.

Identity on ATProto is more portable than Mastodon: users have DIDs (decentralized identifiers) that theoretically allow migration between servers without losing followers. Account migration works in practice. But the PLC (Public Ledger of Credentials) directory, a centralized service that stores PLC operation logs and serves DID documents including service endpoints and key material, remains controlled by Bluesky PBC. For Bluesky-hosted accounts, the company still typically participates in key management and recovery. Bluesky has discussed transferring PLC governance to an independent Swiss association, acknowledging the current centralization.

Bluesky has emphasized portability and credible exit more than fully decentralized operation today: ensuring users and developers could leave if needed, not that the network presently operates without central control. Blacksky is also building a parallel stack beyond the relay, including its own PDS, labeler, custom feeds, and an App View under development, and the Free Our Feeds initiative has announced additional infrastructure intended to follow. The architecture enables competition, and the proof that a second independent full-network relay is operationally sustainable is now on the record. Most of today's traffic still concentrates in the dominant relay, but "only one" no longer describes the situation.

Blockchain-Based: Wrong Tool for Social Coordination

Some social protocols use blockchains, but this reflects a category error about what social communication requires.

Global consensus is unnecessary for social communication. Blockchain's core innovation is solving double-spending through global consensus, meaning every participant must agree on a single transaction history, a guarantee critical for money because the same coin cannot be spent twice. Social posts have no such constraint. A message can exist on multiple servers without creating inconsistency. There is nothing to "double-spend."

Blockchain architecture also forces data replication, requiring all participants to download all data. Every node stores the complete history. For monetary transactions this ensures no one can cheat; for social posts it means every user must store every other user's content. This scales catastrophically and serves no purpose.

The global consensus model creates full metadata exposure. Every transaction and follow relationship is visible to every participant. For monetary systems this transparency enables verification; for social systems it creates surveillance at total scale. Users cannot selectively share with trusted parties when the architecture demands universal broadcast.

Finally, using blockchain typically requires tokens, adding financial friction to communication and inviting speculation that distorts usage.

Nostr inverts these properties. Users store events only on relays they control or trust, and no requirement exists that other users see those events. Alice can post to her personal relay without broadcasting to the world, and this selective visibility is precisely what social communication needs and what blockchain architecture prevents.

Why Nostr's Simplicity Provides Advantages

Nostr's entire protocol surface is a signed JSON file and a WebSocket connection to a relay. A developer can implement a working client in a weekend from a spec that fits in a single document. ActivityPub and ATProto each impose complex federation logic or large-scale firehose indexing infrastructure; blockchain-based protocols add full participation in a consensus layer on top of that. Nostr imposes none of these, which is why the client network is wide, with thousands of independent implementations competing on design and target audience, each able to reach the full network without special access.

Relay scaling is also independent. Each relay stores only what it chooses to store and serves only its own users. A relay does not need to know about or coordinate with other relays to function. A small personal relay serving fifty users runs on a cheap VPS and handles

its load without concern for what a large relay serving a million users is doing. The resource cost scales with the audience served, not with the size of the total network, which is why the barrier to running infrastructure remains low enough for individuals and hobbyists.

Chapter Summary

Nostr addresses the identity-capture problem through protocol design. Identity is a user-controlled keypair, content is signed and distributed through relays the user chooses, and exit remains possible without losing identity. Anyone can operate a relay, and moderation happens at the relay and client layers through market services instead of protocol-level authority. Reputation emerges through web-of-trust and domain-based verification instead of platform checkmarks. The signed-event architecture is simple enough that the same infrastructure carries long-form articles, classified listings, live video, wiki entries, audio rooms, software releases, and encrypted group messages alongside short social posts. One keypair can serve many purposes and carry reputation across contexts, and the NIP system allows protocol evolution without central authority. Alternatives fall short for different reasons: Mastodon keeps identity server-dependent, Bluesky's reference architecture creates heavy relay and indexing requirements that still concentrate much of today's network even as portability improves, and blockchain-based solutions impose global consensus requirements unnecessary for social communication.

Privacy limitations are real. Nostr provides pseudonymity, not anonymity. Relay operators see metadata and the social graph is largely public; persistent keys accumulate behavioral patterns that combine into identifiable profiles over time. Key management remains unsolved: key loss is permanent and compromise is irreversible, while rotation conventions remain inconsistent across clients. These problems are inherent to self-sovereign identity systems and do not yet have mature solutions. The Marmot protocol addresses the encrypted-group-messaging and metadata-hiding problems directly by composing MLS with Nostr's relay network. The

protocol itself optimizes for censorship resistance and user control; users needing strong privacy still supplement it with anonymization tools or purpose-built private systems.

Current centralization tendencies exist despite the decentralized design. A few large relays carry most traffic and a few clients attract most users; content discovery routes through a handful of indexing services, reproducing the concentration pattern the protocol was designed to prevent at the identity layer. Whether market competition maintains enough alternatives to prevent reconcentration is an empirical question the protocol alone cannot answer. Nostr shows at the protocol level that complex coordination can emerge from simple infrastructure without central control, and that users can retain network effects with less platform lock-in while holding identities that remain both self-sovereign and socially useful. That pattern generalizes beyond social media.

Notes

[1]fiatjaf, "nostr," available at `https://fiatjaf.com/nostr.html`. Protocol specification at `https://github.com/nostr-protocol/nostr`.

[2]Pew Research Center, "Social Media Use in 2025," `https://www.pewresearch.org/internet/`. The 84% YouTube and 71% Facebook figures are from Pew's annual social media tracking surveys. Meta platforms with more than one billion monthly active users as of 2025: Facebook, Instagram, WhatsApp, and Messenger; see Meta Platforms, Inc., quarterly earnings reports, `https://investor.fb.com/`.

[3]Matt Taibbi et al., "The Twitter Files," published as a series of threads on Twitter/X, December 2022 through March 2023. Key installments documented the FBI-Twitter request portal (Taibbi, December 2, 2022), the visibility filtering systems (Bari Weiss, December 8, 2022), and U.S. Central Command's whitelisted influence-operation accounts (Lee Fang, December 20, 2022). A master-thread index was maintained at *Racket News* (Matt Taibbi, December 2022–March 2023).

[4]The coordinated August 5–6, 2018 removal of Alex Jones and InfoWars content from Apple Podcasts, Facebook, YouTube, and Spotify occurred within a 12-hour window with no prior public coordination. YouTube's removal terminated a channel with approximately 2.4 million subscribers; Apple removed five of six Jones podcast feeds from iTunes; Spotify removed all InfoWars content. Contemporary reporting: Oliver Darcy, "Apple, YouTube, Facebook and Spotify take action against Alex Jones and InfoWars in a single day," *CNN Business*, August 6, 2018. Jones retained Twitter/X access until September 6, 2018; see Twitter's statement reported in Casey Newton, "Twitter bans Alex Jones and

InfoWars," *The Verge*, September 6, 2018. The episode is the clearest documented case of multi-platform simultaneous de-platforming predating any formal inter-platform content-moderation agreements.

[5]Amazon Web Services terminated Parler's hosting contract effective January 10, 2021, citing terms-of-service violations; the Parler application and website went offline hours after the termination. Apple and Google had removed the Parler app from their respective stores on January 8–9, 2021. For AWS's stated rationale, see the AWS email to Parler reproduced in reporting: Lauren Feiner, "Amazon says Parler hasn't shown it's able to remove violent content after app store bans," *CNBC*, January 9, 2021. Twitter permanently suspended President Trump's account on January 8, 2021 (Twitter Blog, "Permanent suspension of @realDonaldTrump," January 8, 2021) and removed over 70,000 accounts associated with QAnon coordination; see Twitter Safety, "Removal of QAnon accounts," Twitter Blog, January 11, 2021. The Parler shutdown is the primary documented case of infrastructure-level de-platforming erasing an entire platform's user data in a single action.

[6]Regulation (EU) 2016/679 (General Data Protection Regulation), Article 20: "Right to data portability," `https://gdpr-info.eu/art-20-gdpr/`. Article 20 grants data subjects the right to receive their personal data in a structured, commonly used, machine-readable format and to transmit that data to another controller. The operative limitation is that the portability right attaches to the data subject's own data; data about third parties (including follower contact information) belongs to those third parties and falls outside the Article 20 right. For the UK equivalent, see UK GDPR Article 20 and the ICO's guidance at `https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/right-to-data-portability/`.

[7]Twitter (now X) announced its new API pricing tiers on February 2, 2023; the Basic tier ($100/month, 10,000 tweets/month read access) and Enterprise tier replaced the previously free v2 Academic Research and standard access. The announcement is documented at `https://developer.twitter.com/en/docs/twitter-api/getting-started/about-twitter-api`. Multiple academic research projects, third-party social analytics tools, and migration clients ceased operation within weeks; see reporting in Casey Newton, "Twitter's new API pricing is a disaster for developers," *Platformer*, February 2, 2023; and the Center for Countering Digital Hate's documentation of disruption to their research pipeline. The effective rate increase from free to $100/month for basic access to $42,000/month for enterprise-scale read access ran several orders of magnitude above prior pricing.

[8]Nostr protocol specification, available at `https://github.com/nostr-protocol/nostr`. The protocol was first described by fiatjaf in 2020; the repository contains the base specification and all NIPs.

[9]A snapshot of the Nostr client network as of 2026. Mobile: Damus (`https://damus.io`, iOS, one of the original clients), Amethyst

(https://github.com/vitorpamplona/amethyst, Android), Primal (https://primal.net, cross-platform with a caching service and a built-in Lightning wallet), Nos (https://nos.social), Freefrom, 0xChat. Desktop and web: Coracle (https://coracle.social), Snort (https://snort.social), Iris (https://iris.to), Primal (web), Nostrudel (https://nostrudel.ninja), Jumble. Long-form and publishing: Habla (https://habla.news), Yakihonne (https://yakihonne.com), Highlighter, Stemstr (music). Marketplace and commerce: Shopstr, Plebeian Market. Video and streaming: Flare, Zap.Stream, Nostr.Build (media hosting). Browser extensions for signing: nos2x (https://github.com/fiatjaf/nos2x), Alby (https://getalby.com) which also handles Lightning. Remote-signer infrastructure for key separation: nsec.app and the bunker protocol (NIP-46). Authoritative directory and index: https://nostrapps.com catalogs clients by platform and category; https://nips.nostr.com indexes the standards each client implements. For protocol-level news and ongoing development, https://nostr.how and the Knowledge Is Golden podcast at https://nostr.fm track the network.

[10]Nostr Implementation Possibilities (NIPs), available at https://github.com/nostr-protocol/nips.

[11]Web-of-trust reputation systems do not treat reputation as ownable property. The framework Chapter 6 develops rules out property rights in reputation, since reputation consists of what other people believe and is not a scarce, rivalrous resource. What the trust graph carries is not a property entitlement but an empirical signal: signed attestations and consistent behavior over time, on which other users form their own assessments. The same principle rules out defamation law as a coherent framework for protecting the signal: false speech is answered by counter-speech and by the reader's own trust calculus, not by state enforcement against the speaker. See Stephan Kinsella, "Defamation as a Type of Intellectual Property," in Jörg Guido Hülsmann and Stephan Kinsella, eds., *A Life in Liberty: Liber Amicorum in Honor of Hans-Hermann Hoppe* (Mises Institute, 2024), cited at Chapter 6, note 11; and Murray N. Rothbard, "Knowledge, True and False," in *The Ethics of Liberty* (NYU Press, 1998), chapter 16.

[12]wot-relay: a Nostr relay implementation that filters incoming events by web-of-trust distance from the operator's follow graph. Source at https://github.com/bitvora/wot-relay, maintained by bitvora. The relay accepts events from accounts within a configurable hop distance (default: 3 hops) of a seed pubkey, dropping events from accounts outside the neighborhood without content evaluation. This is a practical deployment of the WoT spam-filtering principle: social proximity substitutes for content moderation while preserving relay operator autonomy.

[13]Zapstore: a decentralized app store built on Nostr, where app releases are published as signed Nostr events and discoverability is governed by the user's web-of-trust graph. Applications published by accounts within the user's trust graph surface preferentially. Source at https://github.com/zapstore/zapstore; application at

https://zapstore.dev/. The relevant NIPs are NIP-51 (lists for app curation) and a draft NIP for software release events. Zapstore is the primary deployed example of WoT as a software-distribution trust mechanism on Nostr, removing the Apple App Store and Google Play as gatekeepers for users willing to sideload.

[14]NIP-05: Mapping Nostr keys to DNS-based internet identifiers, available at https://github.com/nostr-protocol/nips/blob/master/05.md.

[15]NIP-85: Trusted Assertions, available at https://github.com/nostr-protocol/nips/blob/master/85.md. Defines parameterized replaceable events (kind 30382) published by declared service providers, covering subjects identified by public key, event ID, addressable coordinate, or NIP-73 external reference. Each assertion carries computed reputation values as structured tags. The signer is the algorithm's key, not the individual operator's key, separating reputation production from reputation consumption. The kind 10040 opt-in event allows users to specify which provider keys govern their reputation signals, with optional encryption to keep provider selection private.

[16]NIP-23: Long-form Content, available at https://github.com/nostr-protocol/nips/blob/master/23.md. Defines addressable events for articles and blog posts in Markdown format.

[17]NIP-99: Classified Listings, available at https://github.com/nostr-protocol/nips/blob/master/99.md. Implementations include Shopstr and Plebeian Market.

[18]NIP-53: Live Activities, available at https://github.com/nostr-protocol/nips/blob/master/53.md. zap.stream provides a reference implementation for live video streaming with Lightning integration.

[19]NIP-54: Wiki, available at https://github.com/nostr-protocol/nips/blob/master/54.md. Enables decentralized encyclopedia entries with web of trust curation.

[20]HiveTalk: a Nostr-based audio and video conferencing application using WebRTC with Nostr keypairs for authentication and NIP-57 zaps for tipping; see https://hivetalk.nostr1.com/. Nostr Nests: a Nostr-native audio room platform where participation is authenticated through the user's keypair; see https://nostrnests.com/. Both implement NIP-53 Live Activities (https://github.com/nostr-protocol/nips/blob/master/53.md) for real-time space coordination, with Lightning integration via NIP-57 zaps for tips and access control. These deployments demonstrate that the same keypair and relay infrastructure that carries short notes can authenticate participation in real-time interactive applications without a separate account layer.

[21]NIP-51: Lists, available at https://github.com/nostr-protocol/nips/blob/master/51.md. Defines replaceable event kinds for curated sets including follow lists (kind 3, public), mute lists (kind 10000), and private lists whose content field is encrypted with NIP-44. The encrypted variant of the follow list (kind 10002 relay list is separate; kind 3 with encrypted content is the private follows mechanism) allows

a user to conceal their social graph from relay operators. Client adoption of the encrypted variant is inconsistent as of 2025; most major clients publish follow lists in cleartext as kind 3 events with no encryption.

[22]Cure53, "Audit of NIP-44 Encryption for Nostr," December 2023, available at `https://cure53.de/audit-report_nip44-implementations.pdf`. The audit reviewed the NIP-44 v2 specification covering ChaCha20 symmetric encryption, HMAC-SHA256 message authentication, HKDF key derivation from the ECDH shared point, and the conversation key derivation procedure. Cure53 found no critical or high-severity issues in the cryptographic design. NIP-44 v2 specification: `https://github.com/nostr-protocol/nips/blob/master/44.md`. For the security weaknesses of the predecessor NIP-04 (AES-256-CBC without MAC, non-standard ECDH), see the NIP-44 rationale section and Yuki Kishimoto's analysis at `https://github.com/nostr-protocol/nips/discussions/620`.

[23]NIP-17: Private Direct Messages, available at `https://github.com/nostr-protocol/nips/blob/master/17.md`. Defines kind 14 sealed messages encrypted with NIP-44 and wrapped in NIP-59 gift wraps (kind 1059). NIP-59: Gift Wrap, available at `https://github.com/nostr-protocol/nips/blob/master/59.md`. The gift-wrap construction signs the outer event with an ephemeral keypair and randomizes the `created_at` timestamp by up to 2 days, so neither the sender's identity nor the real message time is visible to the relay. The `p` tag on the kind 1059 event identifies the recipient; kind 10050 relay list events declare where a user's DM inbox relay lives.

[24]The Messaging Layer Security (MLS) Protocol, RFC 9420, Internet Engineering Task Force (IETF), July 2023, available at `https://datatracker.ietf.org/doc/rfc9420/`. For the architecture document, see RFC 9750.

[25]The Marmot Protocol, available at `https://github.com/marmot-protocol/marmot`. Marmot builds on OpenMLS, a Rust implementation of the MLS protocol, to integrate group encryption with Nostr's relay architecture.

[26]Least Authority, "Audit of White Noise Marmot Protocol Review," November 7, 2025; "Audit of the Marmot Development Kit," March 20, 2026; and "Audit of whitenoise-rs," April 7, 2026, all at `https://leastauthority.com/blog/`. For the changes the audits drove in the codebase, see "What the Least Authority Audit Changed," White Noise blog, `https://whitenoise.chat/blog`, with PR references including whitenoise-rs#539, #540, #548, #634, #637, #638, #642, and #665, and marmot#38 ("MIP-00/02/03: align spec with audit findings"). Least Authority's published audit portfolio includes Zcash and the Ethereum 2.0 specifications; see `https://leastauthority.com/security-consulting/published-audits/`. The protocol maintainers continue to label Marmot experimental and MDK alpha pending a period of specification stability following the audits.

[27]Orbot is the Tor Project's official Android and iOS application for routing device traffic through the Tor network, available at `https://guardianproject.info/apps/org.torproject.android/` and `https://orbot.app/`. Tails is a live operating system that routes all outbound connections through Tor by design; see `https://tails.boum.org/`. Several Nostr clients integrate Tor routing directly: Amethyst (Android) supports Tor via Orbot's transparent proxy mode; Primal's server-side relay communication can be routed through Tor by users running system-wide Orbot. For Nostr-specific Tor usage guidance, see the privacy sections of `https://nostr.how/en/get-started`.

[28]Semaphore: see Privacy and Scaling Explorations, "Semaphore Protocol," `https://docs.semaphore.pse.dev/`, and the audited implementation at `https://github.com/semaphore-protocol/semaphore`. Single-server PIR with sub-second latency on practical databases: Alexandra Henzinger, Matthew M. Hong, Henry Corrigan-Gibbs, Sarah Meiklejohn, and Vinod Vaikuntanathan, "One Server for the Price of Two: Simple and Fast Single-Server Private Information Retrieval" (SimplePIR), 32nd USENIX Security Symposium, 2023; and Alex Davidson, Gonçalo Pestana, and Sofía Celi, "FrodoPIR: Simple, Scalable, Single-Server Private Information Retrieval," Proceedings on Privacy Enhancing Technologies, 2023. Two-server practical metadata-hiding messaging: Saba Eskandarian, Henry Corrigan-Gibbs, Matei Zaharia, and Dan Boneh, "Express: Lowering the Cost of Metadata-Hiding Communication with Cryptographic Privacy," USENIX Security Symposium, 2021.

[29]Mastodon documentation, available at `https://docs.joinmastodon.org/`. ActivityPub is the W3C standard federation protocol at `https://www.w3.org/TR/activitypub/`.

[30]Bluesky and the AT Protocol, available at `https://atproto.com/`.

[31]Bluesky Protocol Team, "Protocol Check-in (Fall 2025)," atproto.com blog, October 20, 2025, confirms that "Blacksky currently runs a full-network relay (using the rsky-relay implementation!)" and is building a full-network App View. For the Blacksky stack, see `https://github.com/blacksky-algorithms/rsky` and `https://docs.blacksky.community`. On planned additional infrastructure, see the Free Our Feeds and Eurosky projects at `https://freeourfeeds.com`. For the broader decentralization trajectory, Bluesky PBC is establishing an independent Swiss association to operate the PLC identity directory (atproto.com/blog/plc-directory-org, September 2025), and the IETF ATP working group was chartered in April 2026.

PART VI

PRAXIS

22
OPERATIONAL SECURITY

> "Only amateurs attack machines; professionals target people."
>
> Bruce Schneier, *Crypto-Gram Newsletter* (2000)[1]

Technical tools fail if humans fail. Encryption protecting your messages is worthless if you post the same content publicly under your real name. Tor's anonymity does not help if you log into your personal accounts through it. Bitcoin's pseudonymity does not protect you if you buy at an exchange that has your identity and then use those coins for sensitive purchases.

Operational security (OPSEC) is the discipline of preventing adversaries from gathering information that could compromise security. It is not a tool but a practice: ongoing attention to the ways human behavior can undermine technical protection.

Read this chapter as a sorting frame and not as a command to adopt every measure at once. Good OPSEC begins by removing the obvious exposures relative to your actual adversary. Chapter 23 turns that logic into staged implementation.[2]

22.1 Threat Modeling: Who Is Your Adversary?

Define Your Specific Adversary

Security is always relative to a specific adversary. Defensive measures are arbitrary until the actor knows whom the defense is meant to resist.

Adversaries differ along two axes: capability and intent. A local network observer on shared WiFi can sniff unencrypted traffic but lacks the resources or motivation to pursue targets beyond the session. A VPN defeats this adversary completely. Passive dragnet surveillance operates at a different scale, capturing traffic in bulk without targeting any individual; encryption and anonymization tools raise the cost of extracting signal from that noise beyond what bulk collection budgets can sustain. Platform surveillance is different again: the services an actor uses voluntarily are themselves the observers, and the defense is choosing services whose business model does not depend on harvesting user data. These three categories share a common feature: the adversary is not looking for you specifically.

Targeted surveillance changes the economics. A corporation actively investigating a specific person can hire expertise and sustain effort across weeks or months, but its resources are finite and its legal authority bounded. A state agency conducting a directed investigation has subpoena power and technical capability that dwarf corporate resources, and it can sustain the effort for years. The cost-benefit calculation shifts at each level: what deters a passive observer does not deter a motivated investigator, and what frustrates a corporate adversary may not slow an intelligence service. Threat modeling exists because the same defensive measure can be overkill against one adversary and dangerously inadequate against another.

Assess Adversary Capabilities and Resources

Adversaries scale from script kiddies running tools they do not understand, through skilled individuals capable of novel attacks, to corporations with hired expertise, law enforcement with compulsory process, and intelligence agencies with both extensive resources and sophisticated capabilities. The resources and sophistication of the actual adversary determine what protective measures are necessary and what are overkill.

Match Defensive Measures to Actual Threats

Defending against NSA when your threat is an abusive ex-partner wastes resources and attention. Defending against a coffee shop hacker when law enforcement is investigating you is dangerously inadequate.

Two common errors are over-engineering and under-engineering. Over-engineering means using Tor to browse recipes when your threat is an advertising tracker; this wastes complexity. Under-engineering means using basic encryption when law enforcement is actively investigating you, creating dangerous inadequacy. A subtler failure is the mismatch: sophisticated technical measures combined with social media over-sharing, where the weak link defeats the strong protection.

Threat modeling asks who might want your information and what resources they would commit to getting it.

Personal Threat Assessment

Your threat profile comes from who you are and how you live. Profession matters: journalists, lawyers, medical professionals, activists, and those handling sensitive information face elevated risks. Jurisdiction shapes both threats and protections, since laws on encryption, speech, and financial privacy vary dramatically. Public profile affects targeting: publishing under your real name or having a history of controversial statements increases visibility. Relationships create interconnected risk: family members' social media can expose your location, and business partners' security practices become your vulnerabilities. Financial situation determines certain threats: wealth attracts different threats than poverty. Political context may elevate risk if your beliefs or activities put you at odds with powerful actors.

Not all information requires equal protection. At the top of any serious inventory are the assets whose loss would do real harm: financial credentials, private communications, medical records, and location patterns that enable physical targeting. Below them sits a tier of exposures that would embarrass but not endanger, and below

that a tier of information that does not matter. Concentrate effort on the first tier and spend it sparingly on the second.

Risk Calibration

Risk can be managed but not eliminated, and the right level of acceptable residual risk depends first on consequence severity, which runs from annoyance through financial loss and legal jeopardy to physical danger. Someone whose worst case is embarrassment calculates differently than someone whose worst case is imprisonment, and the measures that make sense in one case can be either overkill or inadequate in the other. Probability matters too, and here most people make the same systematic mistake: they overestimate the chance of being individually targeted while underestimating how much they are already exposed through routine dragnet collection.

Protection costs include time, money, convenience, and social friction. Measures costing more than the expected harm they prevent are not worth implementing. Sustainability is essential: heroic measures requiring constant vigilance fail when vigilance lapses. Social constraints shape viable options: security measures that isolate you from family or professional networks may cost more than they protect.

Threat models are not static. Reassess when circumstances change: new job, new relationship, changed public profile, political shifts, or after any security incident.

Break the OODA Loop at Observation

Chapter 1 introduced Boyd's OODA loop[3]; Chapter 10 applied it to state surveillance. The same framework guides personal operational security.

Every adversary must cycle through Observe, Orient, Decide, Act. Your first priority is to break the loop at Observe. If the adversary cannot see your activity, later stages become much harder to execute well: orientation degrades and targeting grows less reliable. Prevention of observation is the most cost-effective defense because it can stall the attack chain before serious resources are committed.

When evaluating a practice or tool, ask: does this prevent observation, or does it only complicate later stages? Using encrypted messaging prevents observation of message content. Using a VPN may only complicate attribution after observation has occurred. Both have value, but preventing observation is primary.

If an adversary has already observed your patterns, they have passed the hardest stage. Subsequent stages are easier to execute. This is why operational security failures are often catastrophic: once observation has occurred, the damage compounds through subsequent stages. Handle reuse and metadata leakage are observation failures; pattern correlation extends them into attribution. Each one enables what follows.

22.2 The Weakest Link: Human Factors

Social Engineering Attacks

Social engineering attacks bypass cryptographic protection entirely because they target the one component no protocol can harden: human psychology. The attacker does not break the cipher; the attacker persuades the human to open the door.

Phishing works because humans reflexively trust communications that mimic familiar authority, and a well-crafted fake email from a bank or employer triggers compliance before deliberation begins. Pretexting works because humans are social animals who cooperate with plausible stories; an attacker who constructs a credible scenario ("I'm from IT, we need to verify your account") exploits the target's helpfulness as an attack vector. Baiting exploits curiosity: an infected USB drive left in a parking lot gets plugged in because someone wants to know what is on it. Tailgating exploits courtesy: holding a door open for the person behind you is automatic behavior, and attackers rely on that automaticity to bypass physical access controls.

Each method succeeds by converting a psychological tendency (trust, helpfulness, curiosity, courtesy) into an access path. The pattern is consistent: the attack surface is not the machine but the operator's behavioral defaults. Awareness and procedural discipline

are the only defenses against these attacks, and both degrade under fatigue.

Coercion and Legal Pressure

The $5 wrench attack, examined in Chapter 5, illustrates that physical coercion can compel disclosure regardless of cryptographic strength. A passphrase yields under sufficient physical threat, and no key length changes that calculus. Legal coercion operates similarly: courts can hold individuals in contempt for refusing to disclose passwords, and jurisdictions vary substantially in their rules on compelled disclosure. The United Kingdom's Regulation of Investigatory Powers Act[4] allows courts to order key disclosure with criminal penalties for refusal. In the United States, Fifth Amendment protections apply more narrowly, and courts have split on whether a forced passphrase disclosure constitutes compelled testimony or a physical act.[5]

Technical measures can raise the cost of successful coercion without eliminating it. Deniable encryption systems such as VeraCrypt[6] allow a single encrypted container to have two passwords, each opening a different set of files, so a coerced disclosure reveals plausible content while hiding the sensitive material. Dead-man switches can destroy keys or send alerts when not periodically reset, deterring some forms of detention-based coercion. Threshold schemes distributing key material among multiple parties mean no single person holds everything an adversary needs. These measures shift the attacker's problem from extracting one secret to compromising several people or systems simultaneously, which raises cost without providing a guarantee.

Convenience Shortcuts and Laziness

Security measures that impede convenience get bypassed, and the failure modes are well-worn. Unique passwords for every account are tedious to maintain, so people reuse them and create single points of failure that propagate across services. Signature and checksum verification takes time that people decline to spend,

so unverified software gets installed anyway. Identity separation requires discipline that tired humans skimp on, so streams get crossed. Updates interrupt work, so systems stay unpatched. Each compromise is individually defensible; together they define the gap between the security a system could provide and the security it does.

Sustainable security must account for human laziness by reducing the friction of good practice until it becomes the path of least resistance. A password manager makes unique passwords easier than reuse. A hardware key makes phishing-resistant authentication faster than typing a one-time code. Encrypted messaging apps[7] that match the UX of surveilled alternatives get adopted, while apps requiring manual key exchange stay on the shelf. The discipline compounds only when the secure practice is also the convenient one.

Humans Fail Before Technology Fails

In many security breaches involving good cryptography, the decisive failure is human: using weak passwords, reusing credentials across services, falling for phishing, mixing identities, social media over-sharing, or trusting compromised collaborators. Silk Road's collapse came from handle reuse and server misconfiguration. LulzSec members were caught because they shared information with a collaborator who had already been turned by the FBI. Reality Winner was identified through printer metadata, not a cryptographic break. In each case the technical layer held and the human layer failed.

The pattern has a consistent shape. Attackers who cannot defeat the cryptography probe the humans around it. They socially engineer the administrator, wait for a moment of fatigue, or cultivate a trusted insider. Technical configuration sets the floor; human discipline determines whether the system stays above it. OPSEC is primarily the practice of managing behavior so that the gap between the floor the tools provide and the ceiling human error allows is as small as the operator can make it.

22.3 Technical Security Fundamentals

Device and Operating System

Security starts with the operating system. The OS mediates every interaction between applications and hardware, which means a compromised OS undermines every security measure built on top of it. This is why purpose-built systems exist. Qubes OS[8] isolates applications in separate virtual machines so that a compromised browser cannot reach files or keys in another compartment; the architecture assumes breach of individual components and contains the damage structurally. GrapheneOS[9] hardens Android with improved sandboxing and verified boot, and provides a sandboxed approach to Google Play in place of integrating Google services into the OS, closing attack surface that stock Android leaves open by default. When specialized systems are impractical, the principle still applies: disable telemetry and unnecessary services, run daily work from standard user accounts instead of administrator privileges, and enable features like Lockdown Mode on Apple devices. Each measure reduces the attack surface the OS presents.

Physical access to an unencrypted device is total compromise, and this is the threat that disk encryption addresses. Full-disk encryption (LUKS on Linux, FileVault on macOS, BitLocker on Windows)[10] ensures that a lost or stolen device yields nothing readable without the decryption key. Firmware passwords prevent an attacker from booting unauthorized media, and verified boot on supported hardware ensures the bootloader itself has not been tampered with. Screen locks with short timeouts and strong passwords address the brief-absence threat; biometrics are convenient but can be compelled in some jurisdictions, while PINs and passwords retain stronger legal protection. The same logic of attack-surface reduction applies to software: every installed application and every running network service is a potential entry point, and what is not present cannot be exploited.

Network Security and Key Management

Network security prevents eavesdropping and man-in-the-middle attacks across the layers where they occur. HTTPS should be used everywhere; verify TLS certificates and use browser extensions that enforce HTTPS. Public WiFi should be treated as hostile, so use a VPN on untrusted networks. DNS queries can leak information about what you access even when the traffic itself is encrypted, so use DNS over HTTPS or configure trusted resolvers.[11] Configure the firewall to block incoming connections that are not needed, because every open port is a potential entry point.

Cryptographic keys are the material the whole system rests on, and they must be both protected and recoverable. Keys should be encrypted at rest, preferably with hardware protection such as HSMs or hardware wallets that keep the secret off general-purpose computing devices. Keys without backup are vulnerable to loss, but poorly designed backups expand attack surface; the right approach depends on whether the key is disposable or long-lived. Disposable and session keys can be rotated frequently, limiting exposure if any single key is compromised. Long-lived identity keys involve a different tradeoff because rotation destroys continuity, so they require recovery planning with a tested restoration path before any emergency arises: a dead-man switch for incapacitation and multi-signature custody for the key material.

Software provenance and patching complete the picture. Software should come from legitimate sources whose authenticity can be verified. PGP signatures on downloads prove provenance through the web of trust. Zapstore offers an alternative model: developers sign releases with their Nostr keys, and users discover applications through social-graph endorsements, verifying signatures against public keys whose reputation they can evaluate themselves. Hashes prove integrity through checksum verification, and reproducible builds allow verification that a binary corresponds to the published source. Dependencies can be compromised, so large dependencies deserve audit. Once software is installed, patches must be applied promptly, since known vulnerabilities are actively exploited. Update sources themselves must be verified, because fake update

prompts are a common attack vector. Software that no longer receives updates should be replaced.

Endpoint Compromise: Commercial Spyware and Forensic Extraction

The device in the defender's hand is itself an attack surface. Two classes of endpoint attack are now industrialized and must be modeled explicitly.

Commercial spyware targets the endpoint through remote exploitation, typically via zero-click vulnerabilities in messaging clients or the media parsers they invoke. The attack installs a full remote-access payload that reads messages after decryption, captures screen content, exfiltrates files, and activates microphone and camera on demand. End-to-end encryption of message transport is irrelevant against a compromised endpoint, because the adversary reads the cleartext the same way the user does. The defender's counterplay against commercial spyware is a discipline composed of several compounding habits. Minimize the attack surface that messaging clients and media parsers expose; use platforms that ship security updates promptly and include exploit-mitigation features; enable the platform's hardened mode (such as Lockdown Mode on Apple devices, which disables several attack-surface features in exchange for reduced functionality); periodically restart the device, because some spyware loses persistence across reboots; and treat unexplained battery drain or network activity as potential indicators of compromise and not as ordinary noise. For targets with elevated threat models, the correct baseline is a hardened device reserved for sensitive work and kept separate from the device used for ordinary life, replaced on a fixed schedule.[12]

Forensic extraction targets the endpoint through physical possession. Tools developed for law enforcement exploit boot-chain and lockscreen weaknesses to extract data from seized phones. Unlike commercial spyware, forensic extraction requires physical possession of the device, and it operates at a price point that makes it routine for ordinary criminal investigations even without the device being unlocked. Every arrest is a potential extraction event, and the

extracted data is retained in evidentiary systems whether or not a prosecution follows. The defender's counterplay is tighter. Disk encryption with a high-entropy passphrase sets the cost floor for an extraction attack; the floor is not infinite, but it is high enough that generic extraction tools fail against it. Features that deny the extraction tool its exploitation window matter, including USB restricted mode on GrapheneOS and iOS, which disables the USB data path after the device is locked. Turning the device off before handing it over (or before it is taken) moves the device into a stronger cryptographic state than a locked-but-running device, because cold boot eliminates keys that may otherwise remain in memory. Each of these measures raises the cost and lowers the probability that a given extraction attempt succeeds, and the accumulated friction is the defense.[13]

When the channel is encrypted, the endpoint is attacked: this is what the adversary chapters established and this chapter confirms. Encrypting the channel is not enough; hardening the device must be maintained under the assumption that endpoint attacks are now routine instead of exceptional.

Supply-Chain Integrity: Reproducible Builds and Signed Releases

A privacy tool the user cannot verify is a privacy tool the user trusts by faith. The reproducible-builds project, coordinated across most major Linux distributions since the mid-2010s, addresses one half of the verification problem by making it possible to build a binary from source and confirm that the resulting binary matches the binary the project distributes.[14] Debian reached 95% reproducibility for its main archive by 2024; Arch and NixOS track similar progress; Bitcoin Core has been reproducibly built for years; Tor Browser ships with documented reproducible-build instructions. The other half is signed-release transparency, which PGP, Sigstore and the in-toto attestation framework provide by publishing cryptographic signatures over build provenance in a transparency log any user can audit. Zapstore carries the same idea into the Nostr network, letting developers sign release artifacts with their Nostr keys and letting

users discover and verify those signatures through the social graph instead of through a central authority.

The practical workflow is narrower than the theoretical program. Most users download pre-built binaries and verify the distribution's signature against a key the distribution's package manager already trusts. The reproducible-build property matters even when the user does not re-execute the build, because any party (a security researcher, a community group, an adversarial reviewer) can re-execute the build and raise an alarm if the result diverges. The reproducibility program makes the signing infrastructure falsifiable in public, which is the reason the program was started. The SLSA framework extends the same idea across enterprise supply chains with tiered attestation requirements that enterprise deployers can apply to their software bills of materials. Privacy tools depend on more than open-source licenses; they depend on the reproducibility and attestation infrastructure that makes the source-to-binary mapping auditable, and the infrastructure is how the open-source promise becomes checkable.

Hardware Security Tokens

Dedicated cryptographic hardware falls into two overlapping categories.[15] FIDO2 authentication tokens (YubiKey as the closed-firmware industry standard; Nitrokey and SoloKey as open-firmware alternatives) bind each credential to a specific domain at registration, making phishing-resistant second factors structurally sound. Bitcoin hardware wallets (Coldcard for air-gapped Bitcoin-only signing; Trezor and Ledger for broader multi-asset use) keep the signing key off general-purpose computers. The distinction between the two categories is narrowing as some devices support both, and the appropriate choice depends on which keys require hardware protection.

Open-firmware laptop platforms (Framework, System76, MNT Reform, Purism Librem) let users run coreboot or Heads firmware in place of proprietary UEFI; Heads adds measured boot that detects physical tampering the ordinary firmware stack does not support.

Physical Isolation

Some threat models require explicit physical-isolation measures in addition to the software ones.[16] Faraday bags block cellular, WiFi, Bluetooth, and GPS emissions from phones carried into sensitive contexts, denying the device's presence from producing a subpoenable movement log. RF-blocking wallets defeat passive contactless reads of passports and credit cards. Laptops with hardware kill switches let the user physically disconnect the microphone and wireless radios instead of trusting software to disable them. Consumer routers running OpenWrt replace opaque residential-gateway firmware with an inspectable stack, and pocket-sized OpenWrt travel routers turn an untrusted hotel or airport network into a user-controlled tunnel at the first hop.

Physical isolation complements the cryptographic and compartmentalization measures described earlier; it does not replace them. The case where it becomes structurally necessary is the case where the software stack has been compromised or legally compelled, and the physical boundary is the last resort that denies the adversary the channel no software guarantee can close.

22.4 Compartmentalization and Identity Separation

Never Cross Identity Streams

The most common OPSEC failure is identity crossover: a single point of correlation that connects an anonymous identity to a real one. Compartmentalization works by ensuring that an adversary who observes one identity cannot connect it to another. Any shared element between identities defeats this purpose, and the shared element need not be obvious.

A username reused across contexts gives an adversary a free correlation point; one search links the anonymous persona to the real account. Personal email addresses used to register anonymous accounts tie them at the provider level. Shared browsers carry fingerprinting data (installed fonts, screen resolution, plugin configuration, canvas rendering) that can identify the same machine across sessions even without cookies. IP addresses link identities

at the network level, and ISP logs can confirm that both identities operated from the same connection. Even writing style can serve as a correlator; linguistic analysis of word choice and sentence structure can link texts authored under different names.

The critical property of identity crossover is irreversibility. Once a link between identities exists in an adversary's dataset, it cannot be unlinked. Deleting the username or closing the email account does not erase the correlation from logs already collected. Compartmentalization is a property of the system's entire history, not its current state, which is why it must be maintained from the first action under each identity.

One Identity per Purpose

Each purpose should have its own identity, and purposes should not mix. A real identity is for official matters: employment, family, banking, anything that must connect to a legal name. A pseudonymous professional identity is for work one wants attributed to a consistent persona but not to the legal name behind it. Anonymous identities are for activities where even a consistent pseudonym is undesirable. Mixing purposes within an identity creates links that cannot later be broken, because once a piece of real-world information has touched a pseudonymous persona, the persona is no longer pseudonymous in any meaningful sense.

Hardware Separation Between Identities

Ideal separation uses different hardware for different identities. A laptop used only for anonymous activity cannot leak information to your real identity through device fingerprinting. Different operating systems serve different compartmentalization needs. Tails[17] is a live operating system that boots from USB and routes all traffic through Tor, leaving no trace on the host computer; it is ideal for one-off sensitive activities where you want no persistent state and complete network anonymity. Qubes is a desktop operating system that compartmentalizes different activities in separate virtual machines running simultaneously; it is ideal for ongoing work across

multiple security contexts, where you need to maintain separate identities persistently while switching between them throughout the day. GrapheneOS is a hardened mobile operating system for Pixel phones; it provides strong device security and privacy for mobile computing but is a phone OS, not a desktop solution. The choice depends on use case: Tails for anonymous sessions without persistence, Qubes for compartmentalized daily computing, GrapheneOS for secure mobile. Less ideal but more practical for those who cannot adopt specialized systems: different browsers, different profiles, different user accounts on a standard operating system.

Network and Temporal Separation

Different identities should use different network paths, because network correlation is one of the most reliable deanonymization techniques and the easiest to apply at scale. Home internet connects to a real identity and should stay connected to it. Public WiFi or mobile data purchased anonymously handles anonymous activity. If VPNs are used, different providers should cover different identities. Where the stakes warrant it, the real identity can use ordinary internet while the anonymous identity uses Tor. Serious compartmentalization requires that identities do not share a network address or a traffic pattern visible to the same observer.

Temporal correlation is the other side of the same problem. Activity patterns link identities even when names and addresses do not. If two identities are always active during the same hours, they may be the same person working the same schedule. Rapid response to the same external event across two identities can confirm the link through timing alone. Two identities that both go quiet during the same vacation are a stronger signal than either one active in isolation. Varying activity patterns deliberately and introducing delays where they cost little can defeat the naive form of the attack; nothing short of hard separation of devices and schedules defeats the sophisticated form.

22.5 Organizational Security and the Infiltration Problem

Operational security applies to groups as much as to individuals. A movement or local trading circle can fail through the same kinds of leakage that expose a single person, but groups face one danger that individuals don't: infiltration.

Every organization that matters should assume hostile actors will try to enter it, because the state's record leaves little room for innocence on this point. When authorities decide that a movement, newsroom, activist network, or dissident market matters, they do not rely on passive observation alone; they cultivate informants and place agents where trust is already forming. The right response to this reality is design, not panic. Structural arrangements that limit what any single compromised participant can give away matter more than ad-hoc suspicion of whoever happens to seem out of place at a given moment.

Detection Is Not Enough

Most advice on infiltration starts with detection: watch for background stories that do not fit, for sudden access to resources that an ordinary participant would not have, for the person who always pushes the group toward reckless action. Some of this advice is sound and worth internalizing. None of it solves the main problem, which is that competent infiltrators are trained for exactly these suspicions and will pass them.

A movement that relies on intuition alone will often miss the real informant while accusing the wrong people, and that failure carries a second cost beyond the immediate miscarriage: detection-obsessed groups turn paranoid and collapse from within as mutual suspicion and loyalty tests consume the work itself. The state does not need every informant to produce prosecutions. It is often enough to make the organization consume itself from the inside while the external investigation takes its time.

Vigilance still matters. It is a second line of defense. It cannot be the first.

Design for Damage Containment

The primary defense is structural. Build so that even a successful infiltration does not achieve its purpose.

Cryptography already works this way. A secure system assumes the adversary knows the protocol and the code but not the secret keys. Organizational security needs the same posture. Assume a hostile actor may enter the room, and assume that someone may later be arrested, pressured, or turned. Then ask what that person could expose.

When one participant holds too much information, infiltration pays well. When sensitive knowledge is distributed and high-risk actions are compartmentalized, the value of the mole drops further if private communications are protected end to end. An informant may report what was said in one meeting, but cannot map the whole network or expose plans they were never allowed to see.

Compartmentalize Knowledge, Not Only Devices

The same discipline that applies to devices and networks applies to knowledge within a group.

Not everyone needs the full member list or the infrastructure map, and not everyone needs the supplier chain, wallet structure, meetup locations, or contingency plan. A collaborator should know what their role requires and little beyond it.

Bounded knowledge is a sign of competence, not distrust. Mature organizations do not confuse warmth with exposure, and they understand that trust can be sincere and still bounded.

The same rule applies to communications. Full operational detail should not live in the broadest channel, and no single chat room should become the archive of the whole project. Conversations split by role and time horizon localize any eventual compromise, so that one participant turning hostile or one account falling does not hand an outside party the entire project at once.

Redundancy and Flat Structures

Hierarchies make infiltration profitable because they create obvious high-value targets. Compromise the leader, the treasurer, the relay operator, or the one person who knows how everything fits together, and the damage spreads fast.

Redundancy changes the economics in the opposite direction from bounded knowledge. More than one person should know how to perform any critical function; more than one channel should exist for critical communication; more than one venue should exist for meeting and trade. A reasonably flat structure reinforces the point, so that no single disappearance or defection can freeze the whole network. The two principles sit together: compartmentalize information so the compromise of any one person is survivable, and duplicate capability so the loss of any one person is survivable too.

Groups still need coordination and standards, but in a form that does not turn every key role into a single point of failure.

A serious privacy network assumes that compromise will happen and that the network must still work the next day, instead of assuming that everyone inside it is clean.

22.6 Venue Security and Physical Nodes

Groups that meet in person inherit a second problem. A recurring room or merchant back room is useful because people know where to find each other; the same stability also makes the place easier to watch. Physical nodes need rules of their own.

Access control comes first. A sensitive venue keeps its real function opaque from the street and keeps guests within the spaces their visit requires. Public-facing space should be separated from storage and sensitive conversation space when the use case permits. Guest lists should stay minimal, and someone still needs to know who invited whom and who is responsible for a newcomer. The point is to keep casual curiosity and opportunists from learning the whole layout for free.

Approach paths deserve the same attention as the room. A venue used for trade or coordination should avoid training everyone into

one visible routine. Repeated meeting time, one entry path, house WiFi used by every attendee, and one parking pattern create the same gift for observers: a stable pattern to map. Better practice is plain: vary routes and timing where the cost is low, keep sensitive devices off the venue network. Agree on a signal in advance to warn people off a compromised site. A room that can be replaced is safer than one treated as irreplaceable.[18]

22.7 Surveillance Detection

Prevention is preferable to detection, but detection enables response. Recognizing when you are under surveillance, whether physical or digital, allows you to modify behavior before compromise becomes complete.

Physical Surveillance Indicators

Physical surveillance leaves traces if you know what to observe. The same person appearing in multiple unconnected locations is one of the strongest indicators; one appearance may be coincidence, but repeated appearances deserve attention. Vehicles that appear repeatedly, especially if they contain occupants who do not exit, warrant attention. People who seem to have no purpose, loitering without apparent reason, reading newspapers for extended periods, or making phone calls that never end, may be conducting surveillance. Sudden behavioral changes in your environment, such as new "regular" faces at your usual locations, merit scrutiny.

Detection requires establishing a baseline of normal activity. Know who typically populates your regular environments: the coffee shop, the commute, the neighborhood. Without knowing what is normal, you cannot recognize what has changed.

Counter-surveillance routes test for followers. Vary your routine unpredictably. Use routes that force followers to expose themselves: dead-end streets, sudden reversals, entering and quickly exiting buildings with multiple exits. The goal is not to evade but to detect. If detection confirms surveillance, you can then decide whether to continue, modify behavior, or seek assistance.

Digital Surveillance Indicators

Digital surveillance is harder to detect but leaves its own traces. Unexpected account activity, such as login notifications from unfamiliar locations or devices, indicates potential compromise. Password reset emails you did not request suggest someone is probing your accounts. Devices behaving unusually, running hot when idle, battery draining faster than normal, or network activity when you are not using the device, may indicate compromise.

Canary services can detect certain types of surveillance. A dedicated email account that you never use, checked only occasionally, will show login activity only if someone else has accessed it. Files with unique names placed in cloud storage can be monitored; access by anyone other than you indicates compromise. These "tripwires" do not prevent surveillance but reveal it.

Network monitoring can reveal unexpected connections. Tools that display active network connections show what your device is communicating with. Connections to unfamiliar servers, especially during idle periods, warrant investigation. DNS queries to unexpected domains may indicate malware or monitoring software.

Be aware of legal surveillance indicators as well. Unusual law enforcement interest in your associates or legal process served on your service providers may indicate investigation. Some jurisdictions require notification after certain types of surveillance conclude; absence of such notification does not mean absence of surveillance.

Limitations of Detection

Detection is not foolproof. Sophisticated adversaries conduct surveillance designed to evade it, and the absence of detected surveillance does not prove its absence. It also carries its own risks: visible counter-surveillance signals awareness and can accelerate the adversary's timeline, while paranoia damages relationships and judgment and false positives consume resources real threats would have claimed later. The goal is informed, proportionate response, not certainty or paralysis.

22.8 Common Failure Modes (Case Studies)

Handle Reuse and Identity Crossover

Ross Ulbricht's arrest resulted from operational security failures, not cryptographic weaknesses.[19] The critical errors were handle reuse (the "altoid" username appeared on both anonymous promotional posts and personal accounts) and server misconfiguration. The technology worked; human operational failures created the vulnerabilities. This pattern recurs across cases: operators reuse usernames across contexts or include identifiable metadata in uploaded files.

Trusted Collaborator Betrayal

LulzSec was a hacking collective that breached Sony and the CIA among dozens of other high-profile targets during a 50-day spree in 2011. Hector Monsegur ("Sabu"), one of its leaders, was arrested and immediately began cooperating with the FBI.[20] He continued operating within LulzSec while feeding information to investigators.

Other LulzSec members trusted Sabu and shared information with him that led to their arrests. The technical security of their communications did not matter; they were communicating with an informant. No technical measure protects against a trusted collaborator who has been turned. Trust is irreducible; choose carefully whom you trust and compartmentalize what each person knows, because compromise can happen inside the room.

Printer Steganography: Reality Winner

In 2017, Reality Winner, an NSA contractor, printed a classified document and mailed it to journalists.[21] She was identified and arrested within days. The document itself betrayed her.

Color laser printers embed nearly invisible yellow dots that encode the printer's serial number and the exact date and time of printing,[22] and those dots survived the scan of the leaked page. The NSA therefore knew which printer had produced the document and when. Internal logs showed Winner was one of six people who had

printed that document, and further investigation turned up a contact between her work computer and the journalists. She was sentenced to five years in prison.

The lesson extends well beyond printers. Physical documents carry metadata of several kinds, including printer tracking dots, paper batch information, and handling traces that the paper itself records. Digital documents carry their own set in author fields, revision history, embedded GPS coordinates, and a long list of application-specific artifacts that persist after the author believes they have removed them. The safe assumption is that every document, whether physical or digital, contains information its author did not intend to include, and the safe practice is to strip, reprint, or rewrite deliberately whenever the stakes justify the effort.

Blockchain Analysis: Tracing Bitcoin

Bitcoin's pseudonymity is frequently mistaken for anonymity. In 2019, international law enforcement announced a major operation that resulted in 337 arrests across 38 countries.[23] The investigation relied primarily on blockchain analysis, not on breaking encryption or compromising Tor.

The method was direct. Blockchain analysis firms traced Bitcoin flows from the target through to cryptocurrency exchanges, and the exchanges, required by law to collect identity information, produced records tying the transactions back to specific individuals. The blockchain's permanent and public record, designed for trustless verification, became the prosecution's evidence. Users who had believed Bitcoin provided anonymity discovered that every transaction they had ever made was permanently recorded and attributable once any endpoint in the chain touched an identified service.

The lesson: Bitcoin provides pseudonymity, not anonymity. The base layer blockchain is a permanent public record. Privacy requires additional measures: avoiding identified exchanges, coinjoining, transacting through Lightning Network with appropriate channel management, or using ecash systems that break the transaction graph. Pseudonymity is useful but is not the same as anonymity.

Photo Metadata: EXIF Exposure

In 2012, software entrepreneur John McAfee was evading authorities in Central America. Journalists accompanied him and published photos documenting his flight.[24] One photo contained EXIF metadata including GPS coordinates. Within hours, observers had identified his precise location at a resort in Guatemala. He was arrested days later.

The journalists knew about metadata risks; they had warned each other. The failure occurred during file handling at the publication's headquarters. Someone uploaded the original file; the stripped version never went out. One mistake in a chain of careful handling was sufficient.

Modern smartphones embed extensive metadata in photos by default: GPS coordinates precise to meters, device model, date and time, sometimes camera settings and thumbnails of previous edits. This metadata survives casual inspection; viewing a photo does not reveal the hidden data. Sharing an unstripped photo can reveal your location and your device model down to the serial number. Strip metadata before sharing any image.[25] Better: disable GPS tagging at the camera level for sensitive contexts.

22.9 The Limits of OPSEC

Perfect Operational Security Is Impossible

No one maintains perfect operational security forever. Constant vigilance is exhausting; people slip, and more security measures mean more opportunities for mistakes. Real life creates situations where security must be compromised, and you cannot defend against attacks you do not know exist.

Anyone can make mistakes. Extended operations steadily push the odds of a fatal error upward.

Targeted Sophisticated Adversaries May Succeed

Against a sufficiently motivated and resourced adversary, OPSEC may not be enough. Intelligence agencies have more

resources than individuals, and subpoenas and international cooperation can extend their reach further than any single individual's defenses can stretch. If the adversary can reach the devices or the person, technical measures fail; if they can compromise the hardware or software supply chain, they gain access no downstream discipline can deny them. The limits here are structural, and the right response is to understand where those limits lie, not to imagine the discipline alone will defeat them.

The goal of OPSEC is not to be perfectly secure but to raise the cost of attacking you beyond what adversaries are willing to pay. Against adversaries with unlimited resources, this may not be achievable.

Risk Acceptance as Necessary Component

All activities involve risk, and perfect security is impossible; the serious question is what level of risk is acceptable for the activity in question. The disciplined form of the question has four parts. First, assess the probability of failure and its consequences. Second, identify which measures reduce probability or consequences to an acceptable level. Third, determine what residual risk remains after those measures and whether it is tolerable. Fourth, explicitly acknowledge that some risk remains, which is always more honest than the pretense that mitigation has eliminated it.

Security without risk acceptance is theater. Acknowledging limits enables rational decision-making.

Knowing When to Stop

Diminishing returns apply to OPSEC. Early measures provide large benefits: using encrypted messaging instead of plaintext provides massive security improvement. Later measures provide smaller benefits: using three layers of VPNs instead of two provides marginal improvement. Excessive measures create new risks, as complexity increases the chance of misconfiguration.

At some point, additional measures are not worth their cost in complexity and lost usability. Knowing when to stop is part of good OPSEC.

Chapter Summary

Human behavior is the weakest link in any security system. Silk Road fell through handle reuse and server misconfiguration, not through a failure of Tor or Bitcoin at the protocol layer. LulzSec members were caught because they trusted Sabu, who was already an informant. Reality Winner was identified through printer steganography dots embedded in the document she leaked. Blockchain analysis has traced Bitcoin transactions to identified exchanges and produced hundreds of arrests. John McAfee's location was exposed by GPS coordinates in a photograph's metadata. In each case, operational exposure turned working tools into compromise, the cryptography did not fail on its own. Social engineering, convenience shortcuts, handle reuse, and identity crossover defeat technical protection from the side it cannot defend.

Threat modeling must precede defensive measures. Security is relative to a specific adversary, and the same measure can be overkill against one and dangerously inadequate against another. The OODA framework from Chapter 1 provides the strategic guide: break the adversary's cycle at observation, where prevention is cheapest. Compartmentalization is the operational discipline that enforces the principle, and it is irreversible once broken. Once an adversary links two identities, the correlation cannot be unlinked, which is why different identities require separate handles, hardware, networks, and activity patterns from the first action onward. Groups face infiltration as a structural threat instead of a personnel problem. Organizations must assume hostile actors will attempt entry and design for damage containment through compartmentalized knowledge and redundancy, because detection alone is insufficient against competent infiltration. The same logic extends to recurring physical spaces. Physical meeting spaces survive longer when access

and contingency plans are bounded and venue knowledge does not spread beyond participants who need it.

Perfect OPSEC is not achievable. Fatigue causes slips and complexity creates new failure modes; sufficiently resourced adversaries may succeed regardless of either. Risk acceptance is a necessary component of rational security practice: explicit acknowledgment of residual risk instead of the pretense that mitigation has eliminated it. Operational discipline does not replace cryptographic tools or the institutional structures in Chapters 24 and 25, it is one layer of defense, not the whole system. Knowing when additional measures are not worth their cost is part of good security. Chapter 23 turns the sorting frame into progressive implementation matched to threat models.

Notes

[1]Bruce Schneier, *Cryptogram Newsletter*, October 15, 2000. The observation reflects the reality that technical security often fails through human factors before cryptographic weaknesses enter the picture.

[2]For deep OPSEC reading, Michael Bazzell, *Extreme Privacy: What It Takes to Disappear*, 5th ed. (2023) is the single best applied manual and covers the tactical detail this chapter treats only in principle; Bazzell's podcast *The Privacy, Security, and OSINT Show* accompanies it. Ross Anderson, *Security Engineering*, 3rd ed. (Wiley, 2020) covers OPSEC in the broader context of system design, and Bruce Schneier, *Secrets and Lies* (2000), though older, remains structurally sound on the human-factors side. Edward Snowden, *Permanent Record* (2019) and Andy Greenberg, *Sandworm* (2019) and *Tracers in the Dark* (2022) provide case-study depth. Core open-source tooling referenced throughout this chapter: Qubes OS (qubes-os.org), Tails (tails.net), GrapheneOS (grapheneos.org) on Pixel hardware, KeePassXC (keepassxc.org), Bitwarden (bitwarden.com), YubiKey (yubico.com) hardware tokens, Signal (signal.org), age (age-encryption.org), VeraCrypt (veracrypt.fr), and OnionShare (onionshare.org). The EFF's Surveillance Self-Defense guide (ssd.eff.org) remains the best free resource for specific threat-model walkthroughs.

[3]John Boyd developed the OODA (Observe, Orient, Decide, Act) loop framework through his work as a U.S. Air Force fighter pilot and military strategist. The canonical secondary source is Frans Osinga, *Science, Strategy and War: The Strategic Theory of John Boyd* (Routledge, 2007), which reconstructs Boyd's briefings systematically. Boyd's own briefing documents, including "Destruction and Creation" (1976) and the "Patterns of Conflict" slide deck, are archived at `https://www.dnipogo.org/john-r-boyd/`. The OODA loop's

application to cybersecurity and adversarial decision cycles is developed in Robert Coram, *Boyd: The Fighter Pilot Who Changed the Art of War* (Little, Brown, 2002).

[4]Regulation of Investigatory Powers Act 2000 (RIPA), c. 23 (UK), Part III ("Investigation of Electronic Data Protected by Encryption"), `https://www.legislation.gov.uk/ukpga/2000/23/part/III`. Section 49 creates a power to require disclosure of an encryption key or the information in an intelligible form; Section 53 makes failure to comply a criminal offence carrying up to two years' imprisonment (five years in national security cases). The Investigatory Powers Act 2016, c. 25 (UK), `https://www.legislation.gov.uk/ukpga/2016/25`, updated and extended these powers. For practical analysis, see the Open Rights Group, `https://www.openrightsgroup.org/`.

[5]U.S. courts have reached conflicting conclusions on whether compelling a suspect to provide a passphrase violates the Fifth Amendment's protection against self-incrimination. The leading cases are: *United States v. Fricosu*, 841 F. Supp. 2d 1232 (D. Colo. 2012) (ordering decryption); *In re Grand Jury Subpoena Duces Tecum*, 670 F.3d 1335 (11th Cir. 2012) (Fifth Amendment protection applied); *Commonwealth v. Gelfgatt*, 468 Mass. 512 (2014) (foregone conclusion doctrine applied to compel decryption); *State v. Stahl*, 206 So. 3d 124 (Fla. Dist. Ct. App. 2016) (compelling biometric unlock); and *United States v. Apple MacPro Computer*, 851 F.3d 238 (3d Cir. 2017) (Fifth Amendment protection applied). The EFF maintains a tracker of compelled-decryption cases at `https://www.eff.org/issues/coders/crypto-cases`. For academic treatment, see Orin Kerr and Bruce Brown, "The Law of Encryption," *University of Pennsylvania Law Review* 170 (2022).

[6]VeraCrypt, `https://veracrypt.fr/`, is an open-source disk-encryption tool descended from TrueCrypt (discontinued 2014) and maintained by IDRIX. It supports plausible deniability through hidden volumes: one encrypted container holds two separate encrypted volumes accessed by different passwords, so a coerced disclosure of one password reveals only the decoy files. Source at `https://github.com/veracrypt/VeraCrypt`. The VeraCrypt documentation on hidden volumes is at `https://veracrypt.fr/en/Hidden%20Volume.html`. Independent security audits were conducted by Quarkslab (2016), `https://ostif.org/the-veracrypt-audit-results/`, and NCC Group (2020). For the broader deniable-encryption threat model, see Niels Ferguson, Bruce Schneier, and Tadayoshi Kohno, *Cryptography Engineering* (Wiley, 2010), Chapter 8.

[7]Signal, `https://signal.org/`, is an end-to-end encrypted messaging application developed by the Signal Foundation (non-profit) and Signal Messenger LLC. It uses the Signal Protocol, which combines the Double Ratchet Algorithm with X3DH key agreement to provide forward secrecy and break-in recovery. The protocol specification is at `https://signal.org/docs/`. Signal retains minimal metadata: it does not log message content, contact lists, group memberships, or conversation metadata. For metadata retention, see Signal's

published response to a federal grand jury subpoena (2016), which produced only account registration date and last connection date. Moxie Marlinspike and Trevor Perrin, "The Double Ratchet Algorithm," Signal (2016), https://signal.org/docs/specifications/doubleratchet/. Comparable end-to-end encrypted messengers include Briar (https://briarproject.org/, peer-to-peer over Tor, no central server), SimpleX Chat (https://simplex.chat/, no user identifiers, no phone number required), and Wire (https://wire.com/, open source).

[8]Qubes OS, https://www.qubes-os.org. A security-focused operating system, developed since 2010 by Joanna Rutkowska and the Qubes team, that compartmentalizes user activities into multiple lightweight Xen-based virtual machines with controlled inter-VM communication. Requires VT-x / VT-d hardware support; compatibility list at https://www.qubes-os.org/hcl/. The architecture paper is Joanna Rutkowska and Rafal Wojtczuk, "Qubes OS Architecture" (2010), https://www.qubes-os.org/attachment/wiki/QubesArchitecture/arch-spec-0.3.pdf.

[9]GrapheneOS, https://grapheneos.org. A hardened Android-based operating system supporting recent Google Pixel devices, which the project targets because of their verified-boot and hardware-key-storage support. Features include a hardened memory allocator (hardened_malloc), hardware memory tagging on supported chipsets, and sandboxed Google Play services as an optional unprivileged app. Feature documentation at https://grapheneos.org/features; the project's threat model is at https://grapheneos.org/faq#relationship-google.

[10]LUKS (Linux Unified Key Setup), https://gitlab.com/cryptsetup/cryptsetup, is the standard disk-encryption specification for Linux, maintained as part of the cryptsetup project; it uses dm-crypt as the kernel-level encryption layer. FileVault is Apple macOS's built-in full-disk encryption, documented at https://support.apple.com/guide/mac-help/protect-data-on-your-mac-with-filevault-mh11785/mac; it uses XTS-AES-128 encryption with a 256-bit key. BitLocker is Microsoft Windows's full-disk encryption feature, documented at https://learn.microsoft.com/en-us/windows/security/information-protection/bitlocker/bitlocker-overview; it requires a TPM chip on supported hardware. For a comparative analysis of these implementations and their threat models, see Bruce Schneier, *Practical Cryptography* (Wiley, 2003), and the Arch Linux wiki on disk encryption at https://wiki.archlinux.org/title/Disk_encryption, which covers the technical tradeoffs across platforms.

[11]DNS over HTTPS (DoH) encrypts DNS queries inside HTTPS connections, preventing ISPs and network observers from seeing which domains a user resolves; RFC 8484, https://datatracker.ietf.org/doc/html/rfc8484. DNS over TLS (DoT) provides the same protection over a dedicated TLS connection; RFC 7858, https://datatracker.ietf.org/doc/html/rfc7858. Privacy-respecting

resolvers include Quad9 (https://www.quad9.net/, operated by a Swiss non-profit with a no-logging policy) and Cloudflare 1.1.1.1 (https://1.1.1.1/). For browser-level DoH configuration, see Mozilla's documentation at https://support.mozilla.org/en-US/kb/firefox-dns-over-https. For the broader DNS privacy threat model, see Geoff Huston, "DNS Privacy," APNIC Blog (2019), https://blog.apnic.net/, and the EFF's analysis of DNS surveillance at https://www.eff.org/deeplinks/2019/09/encrypted-dns-could-help-close-biggest-privacy-gap-internet.

[12]Commercial spyware (Pegasus, Predator, Graphite) uses zero-click exploits to compromise messaging endpoints and read post-decryption content; Apple's Lockdown Mode and periodic device restarts are the primary consumer-available defenses; for industry detail see Chapter 12 note 7. Apple Lockdown Mode documentation at https://support.apple.com/guide/iphone/turn-on-lockdown-mode-ipho2a2a00ea0/ios. Citizen Lab, https://citizenlab.ca/, for ongoing deployment tracking. Amnesty International Security Lab forensic methodology at https://www.amnesty.org/en/tech/. For the industry overview (NSO Group, Intellexa, Paragon, Candiru, Executive Order 14093), see Chapter 12 note 7.

[13]Cellebrite and Magnet Forensics (which acquired GrayKey) sell forensic extraction devices deployed at police scale globally; iOS USB Restricted Mode and Android equivalents, strong disk encryption, powered-off state, and hardened distributions such as GrapheneOS raise the extraction cost on the defender side; for industry detail see Chapter 12 note 7. Apple USB Restricted Mode documentation at https://support.apple.com/en-us/HT208857 (settings pathway: Settings → Face ID/Touch ID & Passcode → USB Accessories). Google Android lockdown mode and screen-lock controls at https://support.google.com/android. Cellebrite, https://cellebrite.com/. Magnet Forensics (GrayKey), https://www.magnetforensics.com/. GrapheneOS, https://grapheneos.org/features, for auto-reboot into Before First Access state and USB peripheral restrictions. 2024 Cellebrite capability leaks documented in community security reporting at https://9to5mac.com/ and https://grapheneos.org/articles. Electronic Frontier Foundation surveillance self-defense on device seizure at https://ssd.eff.org/.

[14]The Reproducible Builds project coordinates reproducibility work across Debian, Arch, Fedora, NixOS, and major open-source projects including Bitcoin Core and Tor Browser; Sigstore and in-toto provide the complementary signing-transparency infrastructure; SLSA extends the framework into enterprise supply-chain attestation. Reproducible Builds project, https://reproducible-builds.org/, with Debian status at https://tests.reproducible-builds.org/debian/ and the foundational paper by Gernot Heiser, Gerwin Klein, and Toby Murray, "Binary Analysis Is the Only Reliable Way to Trust a Compiler" (2014). Bitcoin Core

Gitian/Guix builds at https://github.com/bitcoin/bitcoin/tree/master/contrib/guix. Tor Browser reproducible build documentation at https://2019.www.torproject.org/docs/verifying-signatures.html. Sigstore, https://www.sigstore.dev/. In-toto, https://in-toto.io/. SLSA framework, https://slsa.dev/. Zapstore, https://zapstore.dev/, with source at https://github.com/zapstore; the architecture uses Nostr events (NIP-94 file metadata and NIP-51 application lists) to carry signatures over release artifacts so the verifier's trust is in the developer's public key instead of in a platform store's review process. For the underlying threat model that reproducible builds and signed-release transparency address, see Ken Thompson, "Reflections on Trusting Trust," *Communications of the ACM* 27, no. 8 (1984): 761–763, the canonical essay on compiler-level supply-chain attack.

[15]YubiKey, Nitrokey, SoloKey, and OnlyKey cover the FIDO2 authentication-token tier (Nitrokey and SoloKey with open firmware); Coldcard, Trezor Safe, and Ledger cover the Bitcoin hardware-wallet tier, with Coldcard as the Bitcoin-only air-gapped reference and SeedSigner available as a DIY stateless build for users who want to assemble their own. Framework, System76, MNT Reform, and Purism Librem are the open-firmware laptop platforms; coreboot and Heads replace proprietary UEFI on supported hardware. YubiKey, https://www.yubico.com/. Nitrokey, https://www.nitrokey.com/, with open firmware at https://github.com/Nitrokey. SoloKey, https://solokeys.com/, with source at https://github.com/solokeys. OnlyKey, https://onlykey.io/. Coldcard, https://coldcard.com/. SeedSigner, https://seedsigner.com/. Trezor Safe, https://trezor.io/. Ledger, https://www.ledger.com/. Framework, https://frame.work/. System76 Thelio, https://system76.com/desktops/thelio. MNT Reform, https://mntre.com/. Purism Librem, https://puri.sm/. Coreboot, https://www.coreboot.org/. Heads firmware, https://osresearch.net/. Documentation on FIDO2/WebAuthn at https://fidoalliance.org/ and W3C WebAuthn specification at https://www.w3.org/TR/webauthn-3/.

[16]Faraday bags (Silent Pocket, Mission Darkness), RF-blocking wallets, Librem hardware kill switches, Turris Omnia and OpenWrt home routers, and GL.iNet travel routers are the physical-isolation stack; the case where they become structurally necessary is the case where the software stack has been compromised or legally compelled. Silent Pocket, https://silent-pocket.com/. Mission Darkness (MOS Equipment), https://mosequipment.com/. OpenWrt, https://openwrt.org/. Turris Omnia and MOX, https://www.turris.com/. GL.iNet, https://www.gl-inet.com/. Purism Librem hardware kill switches documented at https://puri.sm/learn/hardware-kill-switches/. For the broader threat model on physical device seizure and the use of Faraday isolation, see Matt Blaze, "Is Your Computer Listening to You?" *Communications of the ACM* 2015,

and the Freedom of the Press Foundation's protest-grade operational-security guidance at `https://freedom.press/training/`.

[17]Tails (The Amnesic Incognito Live System), `https://tails.net`. A live Debian-based operating system that boots from a USB stick, routes all network traffic through Tor, and leaves no trace on the host computer between sessions. Originated in 2009 as an evolution of Incognito Live CD and is maintained by the Tails project with funding from the Freedom of the Press Foundation and others. Security documentation at `https://tails.net/doc/about/warnings/`.

[18]On venue security, security culture, and bounded physical nodes inside a wider hostile environment, see Smuggler and XYZ, *The Second Realm: Book on Strategy* (Liberty Under Attack Publications, year unverified); Kyle Rearden and Shane Radliff, *Just Below the Surface: A Guide to Security Culture* (Independently Published, 2019); and Shane Radliff and Tom Marshall, *Going Mobile* (*#VanLife Zine*) (CreateSpace Independent Publishing Platform, 2019). These are movement-strategy and security-culture texts, not substitutes for the technical and legal references elsewhere in this chapter.

[19]For analysis of Ulbricht's arrest and OPSEC failures, see Andy Greenberg, "How the Feds Took Down the Dread Pirate Roberts," *Wired*, November 18, 2013; and court documents from United States v. Ross William Ulbricht, 14-cr-68 (S.D.N.Y.).

[20]On Sabu and LulzSec, see Parmy Olson, *We Are Anonymous: Inside the Hacker World of LulzSec, Anonymous, and the Global Cyber Insurgency* (New York: Little, Brown, 2012).

[21]On Reality Winner and printer steganography, see the FBI affidavit in United States v. Reality Leigh Winner, 1:17-MJ-590 (S.D. Ga. 2017). For analysis of the printer tracking dots, see the Electronic Frontier Foundation's documentation on printer steganography.

[22]The machine-yellow-dot (MYD) printer-steganography system encodes a printer's serial number and the print date and time in a pattern of nearly invisible yellow dots across the page; the encoding is invisible to the naked eye but detectable under blue LED light or by digital image analysis. The EFF reverse-engineered the encoding used by Xerox printers and published the specification at `https://www.eff.org/pages/list-printers-which-do-or-do-not-display-tracking-dots`. Edward Snowden warned journalists about this tracking mechanism before the Reality Winner case; the FBI affidavit in United States v. Reality Leigh Winner, 1:17-MJ-590 (S.D. Ga. 2017) describes how the dots were used to narrow the pool of suspects to six and then to one. For the broader landscape of unintentional document identifiers (paper batch codes, toner chemical signatures, ink-jet drop-pattern fingerprints), see Mikko Hyppönen's documentation and the academic work of Edward Delp and colleagues at Purdue on digital watermarking, `https://www.ece.purdue.edu/~ipl/`.

[23]U.S. Department of Justice, "South Korean National and Hundreds of Others Charged Worldwide in the Takedown of the Largest Darknet Child Pornography

Website," Press Release, October 16, 2019. Chainalysis provided blockchain analysis tools used in the investigation.

[24]On the McAfee EXIF metadata incident, see Graham Cluley, "Fugitive John McAfee's location revealed by photo meta-data screw-up," December 3, 2012. The photo's GPS coordinates placed McAfee at a specific resort in Guatemala.

[25]Tools for stripping EXIF and other embedded metadata from image and document files include ExifTool (Phil Harvey, `https://exiftool.org/`, the reference command-line tool supporting hundreds of formats), MAT2 (Metadata Anonymisation Toolkit 2, `https://0xacab.org/jvoisin/mat2`, the tool used by Tails and recommended by the Freedom of the Press Foundation), and the GNOME document viewer's export-without-metadata function. For online verification, Jeffrey's Exif Viewer at `https://exifdata.com/` and the metadata viewer at `https://www.metadata2go.com/` allow inspection of uploaded files. The Freedom of the Press Foundation's SecureDrop documentation, `https://docs.securedrop.org/`, covers metadata removal as part of its source-protection workflow. For the threat model, see Harlo Holmes, "Metadata: The Biggest Little Privacy Problem," Freedom of the Press Foundation, 2014.

23
IMPLEMENTATION STRATEGY

> "The world in which cryptography is practiced is no longer the world in which cryptography was designed."
>
> Moxie Marlinspike, "The Network in Motion" (2016)[1]

Privacy implementation is not a single decision but an ongoing process. It requires an accurate picture of current circumstances and clear understanding of personal risks, alongside progressive skill development within community. The goal is not perfection but improvement: moving from wherever you are toward greater autonomy.

23.1 Starting Where You Are

Honest Assessment of Current Exposure

Effective privacy implementation begins with an accurate picture of current circumstances. Most people, upon reflection, discover they have exposed more information than they realized.

Consider what a determined investigator could discover from public information alone. A digital footprint includes social media profiles, forum posts, public records, data-broker aggregations, news mentions, and professional directories. Financial records

reveal property ownership, court filings, political donations, and business registrations. Behavioral patterns emerge from social-media check-ins, tagged associates, public follows and likes, and schedule traces visible in posting times. Historical exposure compounds the problem: material posted years ago remains accessible in archives, and reused usernames create correlation opportunities that no single deletion can close.

This assessment should be uncomfortable. The goal is not self-criticism but clear-eyed understanding of the starting point. Privacy improvement requires knowing what has already been exposed.

What Can and Cannot Be Changed

Some exposure is permanent. Information once public cannot be fully retracted, and data already collected by corporations and governments cannot be reclaimed; patterns already established leave a trail that no later deletion reaches.

But ongoing exposure can be modified. Controllable factors include future posting behavior, service choices, communication methods, and financial tools. Partially controllable factors include professional requirements (which may require some disclosure) and family connections (others' posting affects you); legacy accounts can be deleted but their archived copies cannot be reached. Largely uncontrollable factors include government records and data already sold to brokers; historical posts sit in archives no deletion request reaches. Effective strategy focuses energy on what can be changed while accepting what cannot. Obsessing over past exposure wastes resources better spent improving future practices.

Incremental Improvement Over Perfection

Perfect privacy is unachievable. Pursuing perfection leads to paralysis or burnout. The practical goal is continuous improvement: making your position better today than yesterday, better this year than last year.

This requires accepting imperfection, since every system has vulnerabilities and every practice has limitations; the question is not

whether vulnerabilities exist but whether your position is improving. Prioritize high-impact changes: some modifications provide large privacy improvements with minimal effort, while others require significant effort for marginal improvement. Build sustainable habits, because practices that require constant vigilance fail when vigilance lapses; sustainable privacy comes from habits that become automatic, not heroic efforts that cannot be maintained. Avoid all-or-nothing thinking: partial implementation provides partial protection. Using encrypted messaging with some contacts while using SMS with others is better than using SMS with everyone. Imperfect progress beats perfect paralysis.

23.2 Before You Begin: Apply Your Threat Model

Before selecting implementation steps, apply the threat modeling framework from Chapter 22 to your specific circumstances.[2] Your profession, jurisdiction, public profile, relationships, financial situation, and political context all shape which measures are appropriate and which are unnecessary. Identify what information would cause serious harm if exposed versus what would only embarrass. Concentrate resources on protecting the former; accept that protecting everything equally means protecting nothing effectively.

This assessment should precede tool selection. Generic advice assumes generic threats, but your situation is specific. The implementation steps that follow are organized progressively, but which level is appropriate for you depends on your threat model, not on a universal standard.

A rough rule helps. Readers mainly exposed to ordinary data extraction or account compromise should start with the foundation level. Readers trying to reduce financial surveillance or platform dependence usually need the intermediate layer as well. When the threat includes targeted investigation or serious identity separation, the advanced layer stops being optional.

23.3 Progressive Steps: From Simple to Advanced

These levels are thresholds, not badges of seriousness. Stop when your threat model is covered and the practice is sustainable.[3]

Beginner Level: Foundation Building

The foundation level addresses the adversaries most people face: automated credential theft and passive data harvesting by platforms and data brokers. Each measure below follows from a specific threat, not from a generic recommendation to "be more secure."

Password reuse is the single most common authentication vulnerability because a breach at one service hands the attacker valid credentials for every other service sharing the same password. A password manager eliminates this class of attack by generating unique credentials per account, reducing the blast radius of any single breach to that service alone.[4] Two-factor authentication addresses the next failure mode: stolen credentials that work even when unique. Hardware security keys using WebAuthn or FIDO2 resist phishing because each credential is cryptographically bound at registration to the specific domain that issued it; the browser enforces that scope at sign-in, so a look-alike phishing page cannot obtain a signature the real service will accept. Authenticator app codes are not phishing-resistant in the same sense, a real-time relay page can capture the code and forward it to the real service before it expires, but they still help against bulk credential-stuffing. SMS codes are weakest because they add SIM-swap[5] and carrier-interception risk on top of the same relay vulnerability.

Unencrypted communications are observable by every intermediary that handles them. Signal provides end-to-end encryption with minimal metadata retention. Email is structurally not private; messages traverse multiple servers in plaintext unless both sender and receiver use PGP or S/MIME, and even then metadata remains exposed. Privacy-respecting providers reduce server-side exposure for non-professional communication.[6]

Device encryption addresses the physical-access threat: a lost or stolen device without disk encryption yields its entire contents to

whoever holds it. A privacy-focused browser reduces tracking surface during ordinary browsing. Production-grade privacy browsers now include Mullvad Browser (Tor Browser's anti-fingerprinting work repackaged for use over any network) and LibreWolf (Firefox with the arkenfox hardening profile applied by default); each makes the browser fingerprint indistinguishable from other users of the same browser, which is the structural guarantee that tracker-blocking alone does not provide.[7]

Email addresses serve as universal tracking identifiers across services. Email aliasing services issue a different forwarding address per service, so a breach at any one service cannot be cross-correlated across services, and self-hostable options exist for users who want the forwarding infrastructure to run on their own hardware.[8] These measures address the most common vulnerabilities with minimal lifestyle change; most people can implement them in a weekend.

Intermediate Level: Expanding Protection

The intermediate level addresses a different class of adversary: financial surveillance infrastructure and platform-level data aggregation. Foundation measures protect against opportunistic attackers; these measures begin to resist systemic observers.

Chapter 10 established that financial surveillance operates primarily through choke points where identity is attached to transactions. Bitcoin obtained through full-KYC exchanges reproduces this vulnerability in a new medium.[9] Peer-to-peer exchanges and earning Bitcoin directly for goods or services reduce the identity linkage at the point of acquisition; ATMs in some jurisdictions offer lighter verification requirements, but many now require identity verification for small amounts, so local practice must be checked before assuming privacy.[10]

Spending is a second surveillance surface, and Lightning addresses part of it. Payments route through off-chain channels using onion routing, so individual transactions are not written to the base blockchain and each intermediate node sees only its adjacent hops; this removes the class of attacks that depend on the public transaction graph, including address-reuse clustering and change

detection. The privacy is partial, not absolute: channel opens and closes still settle on-chain as UTXOs that chain analytics can study, the public routing topology combined with timing and amount data can leak information when a wallet holds few channels, and wallets that route through a single provider's infrastructure expose their payment flow to that provider. A self-custodial client connected to a user-run node (Zeus is one option) preserves more of the privacy that provider-dependent setups like Phoenix give up, at the cost of running the node yourself.[11]

A VPN (Mullvad, IVPN) addresses the network-level observer by encrypting traffic between the device and the VPN server, which prevents an ISP or local network operator from inspecting browsing activity. The limitation is structural: the VPN shifts trust from one intermediary to another and does not eliminate observation, only relocate it. Understanding this limitation is part of the threat model, not an argument against using a VPN.[12]

The human-targeted attacks analyzed in Chapter 22 (phishing, pretexting, social engineering) become more relevant as an actor's privacy posture improves, because the human becomes the cheapest remaining attack surface. Security awareness at this stage becomes the primary defense against the attacks that technical measures cannot address, and treating it as supplementary leaves the largest remaining hole. Service migration (replacing privacy-hostile platforms with privacy-respecting alternatives) and identity separation (distinct email addresses and accounts for distinct purposes) follow the same logic: each reduces the number of correlation points an adversary can exploit.

Advanced Level: Stronger Protection

For elevated threat models, stronger measures become appropriate. This level requires significant technical investment but provides protection lower levels cannot achieve.

For network anonymity, Nym aims at stronger protection against timing analysis than Tor's low-latency design by using a mixnet architecture that batches and reorders traffic. Tor remains the more mature and widely used option, with larger anonymity sets and

stronger day-to-day tooling; use Tor Browser for web anonymity. Logging into a personal account through Tor does not hide you from the service itself, which already knows who you are, but it still shields the session from your ISP and any observer sitting between you and the service, so the choice depends on which adversary you are trying to evade. The discipline that matters is identity separation: do not mix anonymous browsing and personal logins in the same Tor Browser session, and use New Identity or a fresh profile when switching contexts, so a network observer cannot correlate the two through a shared circuit. For maximum protection, Tails boots from USB, routes all traffic through Tor, and leaves no trace on the host computer.[13]

Qubes OS isolates applications in separate virtual machines: browser in one VM, email in another, each unable to access the others if compromised. Hardware compatibility is constrained (VT-x/VT-d required, check the compatibility list), but Qubes provides stronger isolation than any single-OS approach. Whonix[14] runs two virtual machines (a gateway routing all traffic through Tor and a workstation isolated from the physical network) and can run inside Qubes for combined compartmentalization and anonymity. GrapheneOS offers hardened Android on Pixel devices with verified boot and hardened memory allocation, plus optional sandboxed Google Play for users who need specific apps without giving Google system-level privileges.[15]

For serious compartmentalization, maintain separate devices for separate identities purchased with cash and used only from locations unlinked to you. Different identities should use different network paths: home internet for real identity, public WiFi or Tor for anonymous activities, with different VPN providers if VPNs serve different purposes.

Self-hosting (email via Mail-in-a-Box, files via Nextcloud) provides maximum control but requires ongoing maintenance. Only proceed if you have the technical capability and time commitment.[16]

Physical security underpins digital privacy. Use firmware passwords, enable verified boot, never leave devices unattended in ad-

versarial environments. For high-threat situations, purchase devices through unpredictable channels to avoid supply chain compromise.

Jurisdiction is a structural variable. Some jurisdictions protect against compelled disclosure, while others do not. Structure activities to benefit from protective legal frameworks.

The Sovereign-Host Tier

Beyond the advanced-user tier lies a narrower category of practice that relocates infrastructure from the user's relationship with commercial providers into the user's own household. The sovereign-host tier is the category the book has been describing architecturally in every chapter on self-hosted systems, and it has matured into a practical configuration during the last several years.

The primary components are ordinary hardware running specific software stacks. A small server (often a single-board computer or a low-power miniature PC) runs a personal-server distribution such as Start9 or Umbrel, each of which packages dozens of applications with defaults that prefer privacy over convenience. An old laptop with a spare SSD works equally well for the same workload, and many first deployments begin on hardware that was headed for a drawer or a recycling bin. A Bitcoin full node runs on the same hardware, so the user's wallet queries their own node and not a commercial provider. A Nostr relay runs alongside the node, so the user's social traffic does not transit someone else's relay unless they choose to cross-post. Storage is local; backups are encrypted with keys the user controls before leaving the box, so external copies can live on cloud storage or a friend's drive without the holder being able to read them; the network path is over residential internet with Tor or I2P[17] available for specific services.

The architectural property of the sovereign host is that it removes the triangular-intervention pathway described in Chapter 10. A commercial provider can be compelled to surveil its customers; a sovereign host answers only to the physical control of its owner. The compulsion pathway that the regulatory architecture assumes still exists for the user's residential internet service provider, but the data flowing through that pipe is encrypted, and the traffic

patterns visible to the ISP reveal that the user is self-hosting instead of disclosing what the user hosts.

The cost structure has converged to make the tier accessible. A small server sufficient to run a full Bitcoin node and a Nostr relay costs roughly the price of a mid-range smartphone, and a repurposed old laptop costs nothing at all. The software stacks are free and maintained by active communities. The maintenance burden is real (security updates and backup discipline) but fits within a few hours per month for a user with baseline technical comfort. The tier has moved from professional systems administration into the practice of a committed subset of the intermediate tier.

What the sovereign host does not provide by default is location anonymity. The host sits at the user's residential address, and its IP address identifies the household to anyone who can correlate IP to location through the standard commercial geolocation databases. A user in an elevated-threat situation can close that gap by exposing the host's services as Tor hidden services or by fronting the connection with a trusted VPN so that incoming queries never see the residential IP directly. Absent either measure, the architecture's value is still real for the ordinary citizen denying the commercial surveillance apparatus the routine traffic it would otherwise harvest. The threat model is lower than the advanced tier's anonymity threat model, and the investment is proportionally smaller.[18]

Local AI Inference on the Sovereign Host

Large language models have become another category of infrastructure the sovereign host can absorb. A prompt sent to a cloud-hosted model is a revealed preference handed to the operator in plaintext, and the operator's terms of service and retention policies determine what happens to it afterward. A prompt processed by a locally run model never leaves the user's device, so no terms of service govern what the operator does with it because there is no operator. Medical, financial, legal, and personal queries that users would be reluctant to hand to a hosted service can be run locally without disclosing them to anyone.

The tooling has matured over the last three years. llama.cpp is the reference C++ inference engine that runs quantized open-weight models on commodity hardware including consumer laptops and mobile devices; the GGUF quantization format is the de facto interchange for open-weight models. Ollama wraps llama.cpp in a developer-facing interface that handles model download and local API serving; LM Studio and GPT4All provide graphical frontends for non-developer users. Apple's MLX framework targets Apple Silicon natively, and Apple's Foundation Models API exposes on-device 3B-parameter models to third-party applications as of the 2025 OS releases. Nvidia's NIM provides containerized models for self-hosted enterprise deployment. None of these are frontier-capability models in the sense that the largest hosted models are, and the gap between frontier and local narrows each year.[19]

The tradeoff the category imposes on the user is capability against privacy. A frontier-class model hosted by a cloud provider produces better answers for many tasks than the largest locally deployable model, and the user who accepts the cloud tradeoff has also accepted the data flow the tradeoff entails. The user who declines the tradeoff accepts a reduction in frontier capability and receives, in exchange, a structural guarantee: no remote party sees the prompt, because nothing leaves the device. For most privacy-relevant queries (medical, financial, personal), a locally deployable model is capable enough, and the privacy guarantee holds without depending on any provider's conduct. The sovereign host is where the category lives in practice: the same hardware that runs the Bitcoin node and Nostr relay also runs the local model, and the inference workload composes naturally with the other self-hosted services on the box.

What Progressive Implementation Accomplishes

Progressive implementation serves multiple purposes. Each level builds capabilities for the next: password manager usage prepares for key management, and basic encryption prepares for more sophisticated cryptographic tools. Gradual change allows new practices to become habits before adding more, since attempting everything at once leads to abandonment. Starting with high-impact

measures ensures that limited resources address the most significant vulnerabilities first. Each step teaches something; practical experience reveals which threats matter and which tools are sustainable for you.[20]

23.4 Finding Your Community

The Role of Community

Privacy is often framed as individual practice, but community expands what individuals can achieve. Tool adoption requires counterparties: encrypted messaging requires others using compatible tools, and Bitcoin transactions require others accepting Bitcoin. Network effects mean that tool utility increases with adoption in your network. Community members share knowledge about tools and emerging threats; learning from others' experience accelerates skill development and helps avoid common mistakes. Technical problems and adversarial situations are easier with community support; isolation makes privacy practice harder to sustain. Some privacy practices require trusted counterparties, since peer-to-peer trading and collaborative projects both require people you can trust.

Local Bitcoin and Privacy Communities

Local communities provide face-to-face interaction that builds trust more effectively than online interaction alone. Many cities have regular Bitcoin meetups that provide introduction to the community and potential trading relationships; quality varies, so visit several if options exist. Maker spaces and hacker spaces often overlap with privacy-focused communities, providing learning environments and social connections. Bitcoin and privacy conferences offer intensive community exposure, though quality varies and research before attending is advisable. Community can also develop organically from existing relationships: friends or colleagues who share privacy concerns can become mutual support networks.

These communities also function as physical bridgeheads. A meetup is a place where online handles become remembered faces

and where future counterparties learn who shows up. The parallel economy grows faster when digital trust touches physical presence.

A recurring workshop or maker-space meetup often becomes the place where people test tools and make the first small trades, while local trust builds to the point where larger trade becomes thinkable. The boundary between community and commerce is often thin for a reason: repeated exchange is one way communities discover whether they can support real economic life together.[21]

Online Communities

Online communities extend reach beyond geographic constraints. Nostr, the decentralized social protocol, enables pseudonymous participation with censorship resistance; finding privacy-focused communities on Nostr provides connection without platform control.[22] Matrix and Element offer federated chat with encryption, and various privacy-focused rooms exist.[23] IRC remains active in technical and privacy communities. Various forums and discussion boards discuss privacy tools and practices, though quality varies and sources should be evaluated critically.

Online communities carry counterparty risk that local meetups can partly resolve through physical presence. You cannot verify whom you are communicating with online, and the same pseudonymity that protects legitimate participants also protects bad actors. Law enforcement cultivates informants in online spaces as readily as in physical ones, and a community's stated values provide no guarantee about the actual composition of its membership. Trust should develop slowly through observed behavior across many interactions, not through claimed credentials or rapid intimacy. Someone who volunteers information about your identity, presses for meeting arrangements unusually quickly, or asks detailed questions about operational specifics before any track record exists deserves caution, not engagement.

Trust Development Over Time

Trust is earned through consistent behavior over time.[24] The mechanism is simple and slow: repeated low-stakes interactions accumulate a track record, and the track record becomes the evidence on which larger commitments rest. Start with participation in discussions and community activities before any high-stakes interaction. Consistent and reliable behavior over months provides real evidence; inconsistency or unreliability provides real evidence too. Extend trust incrementally as experience warrants and withdraw it when behavior contradicts the pattern. There is no shortcut that produces real trust faster than time does, and the feeling of rapid connection with a new community member is more likely a social engineering tell than evidence of real affinity.

Compartmentalization applies inside trusted communities as much as outside them. A trusted contact learns what they need for the interaction at hand, not everything you are working on. Community membership does not eliminate the risk that a participant is an informant, is under pressure, or will at some point make a mistake that exposes what they know. The discipline of bounded disclosure is not distrust toward the individual; it is recognition that trust and exposure are different questions and should be managed separately.

Building Trading Relationships

Peer-to-peer trading requires trusted counterparties, and reputation is the main filter. Established community members with a public history of honest dealing are lower-risk counterparties than strangers because they have something to lose. Initial trades should be small enough that loss is tolerable; size increases only as a track record develops through repeated successful exchanges. Rushing to a large trade before a track record exists is a common mistake that concentrates all the learning in one expensive event.

Escrow reduces risk for both parties in trades with unknown counterparties: funds sit under shared control until delivery is confirmed, not under one side's unilateral authority. Chapter 24 develops the mechanics. For any documentation that protects a

party, the tradeoff between evidence and privacy exposure must be assessed against the actual threat model. A signed receipt that proves delivery also creates a record; whether that record is a liability depends on who might ever see it.[25] [26]

23.5 From Individual Practice to Shared Practice

Privacy practice starts with individuals because individuals can change their habits first, and it becomes durable only when other people around them begin acting on the same rails. A household that uses encrypted communication, or a local merchant circle that trades through recurring counterparties, has crossed a threshold. Privacy is no longer an isolated preference. It is becoming an economic practice.

Household Coordination

Households need shared rules. One person posting family schedules publicly can undo another person's careful location discipline. One spouse using exchange-linked coins for sensitive purchases can expose the other. One child using the family tablet with real-name accounts can collapse identity separation for everyone who shares the device.

Start with simple agreements. Decide which communication tools the household uses for sensitive matters and which devices are shared versus personal. Establish what requires mutual consent before disclosure. If the household uses Bitcoin, decide which balances are spending balances and which are long-term savings, and decide how backup material is stored.

Small Teams and Business

The same logic applies to teams and small firms. A business that wants some privacy cannot run on the assumption that every employee uses the same identity and the same communications habits for every task; roles need separation, and the public-facing account is not the same thing as the operational key. Supplier

communication is not the same thing as public marketing. Customer support is not the same thing as treasury control.

Begin with one line of trade that can clear on cleaner rails. Earn Bitcoin directly where possible. Keep a clear boundary between public reputation and private operations, with treasury control separated from both. Use small counterparties first. Build recurring trade before extending large trust. If a business depends on privacy, it should know which records it needs, where those records live, who can access them, and how much of its activity still depends on bank or platform chokepoints.

One practical entry point is the proxy merchant. In one case the role is an OTC exchanger[27]; in another it is a purchasing agent or a local pickup coordinator. The role localizes exposure. One trusted specialist can handle the boundary and price the risk, sparing ten people ten separate contacts with the surveilled side of the economy. Concentration is the danger, so these roles should stay plural and replaceable.

The First Commercial Loop

The first goal is not scale. The first goal is one closed loop: earn privately, spend privately, then repeat. A designer who takes payment in Bitcoin and pays a contractor in Bitcoin has started that loop. A local merchant who sources from one privacy-conscious supplier and sells to repeat customers through direct messaging has started that loop. A small team has started that loop when it coordinates through encrypted channels and keeps treasury keys separated from daily messaging.

Once the loop works, it can thicken. The counterparty becomes known, the trade repeats; what felt experimental becomes ordinary. A private practice becomes a habit.

23.6 What Not to Do

Do Not Announce Intentions Publicly

A common beginner mistake is publicly announcing privacy intentions. Announcing that you are "going private" attracts attention from those who might not otherwise notice you; adversaries who did not know you existed now know you are trying to hide something. Public announcement establishes a baseline against which changes are measured, making your shift visible where quiet change would be harder to detect. It also creates social pressure to either succeed completely or be seen as failing, pressure that is unnecessary and counterproductive. Announcement often includes information about methods intended, creating a roadmap for adversary countermeasures.

Better approach: Implement quietly. Do not discuss privacy practices with those who do not need to know. If asked, provide minimal information.

Do Not Trust Too Quickly

Community is valuable, but trust must be earned. Informants exist: law enforcement cultivates informants in communities of interest, and enthusiastic newcomers are sometimes not what they appear. Scammers target communities with valuable assets like cryptocurrency or exploitable idealism. Compromised individuals exist: people under legal pressure may cooperate with adversaries to reduce their own exposure, and people who mean well but practice poor security can inadvertently expose those connected to them.

Trust should develop slowly through observed behavior. Extraordinary claims or rapid intimacy are warning signs.

Do Not Over-Complicate Unnecessarily

Complexity is the enemy of security. Complex systems have more failure modes: each additional component is another potential point of failure, and simple systems are easier to secure and maintain. Complex practices are abandoned: practices requiring

constant attention or significant effort get abandoned when life intervenes, so sustainable practices must be maintainable. Using unnecessarily sophisticated tools when simple ones suffice may attract attention; appropriate measures for your threat model are better than maximum measures regardless of threat model. Complex systems also require more mental resources, and limited attention should focus on threats that matter, not theoretical threats addressed by sophisticated measures.

Match measures to actual threats. The adversary you face determines appropriate measures, not the adversary you could theoretically face.

Do Not Underestimate Adversary Capabilities

The opposite error is also dangerous. Sophisticated adversaries exist: intelligence agencies and well-resourced criminals have capabilities beyond common knowledge, so do not assume that commonly known countermeasures address all attacks. Capabilities evolve: what is secure today may not be secure tomorrow, and yesterday's best practice may be today's vulnerability. Unknown vulnerabilities exist, and you cannot defend against attacks you do not know exist; humility about the limits of your knowledge is appropriate. Social engineering bypasses technical measures: the most sophisticated encryption does not protect against an adversary who tricks you into revealing information.

Security requires balance: neither paranoia that prevents action nor complacency that invites compromise.

Common Mistakes to Avoid

Backup discipline is a balance: keys without backup are lost when devices fail, but careless backups expand attack surface. Physical security underpins everything, because digital security means nothing if an adversary can access your devices or observe your screens. Single points of failure require redundancy and compartmentalization. Known vulnerabilities are actively exploited, so security updates must be applied promptly. Metadata neglect

is particularly insidious: encrypted content does not hide who communicates with whom or when, and metadata analysis defeats content-only encryption.

Chapter Summary

Privacy implementation is progressive, not binary. Effective practice builds from foundational measures (password management, two-factor authentication, encrypted messaging, device encryption) through intermediate steps (Bitcoin, Lightning, VPNs, compartmentalization) to advanced practices (Tor, Qubes, GrapheneOS, full identity separation). Each level builds capabilities for the next while providing immediate protection, and which level is appropriate depends on the threat model instead of a universal standard. Beyond the advanced tier sits the sovereign-host tier: a small server or repurposed old laptop running Start9 or Umbrel that hosts a Bitcoin full node and Nostr relay alongside a full suite of self-hosted applications at the user's residential connection. Hardware cost falls from a mid-range-smartphone price to zero when the platform is a laptop headed for a drawer. The tier removes the triangular-intervention pathway that compels commercial providers to surveil their customers, because a sovereign host answers only to the physical control of its owner. The same hardware can now run local AI inference, so prompts medical, legal, or personal enough that users would not hand them to a cloud provider never leave the device at all. Location anonymity is not what the tier offers; denial of the commercial surveillance apparatus's routine traffic harvest is.

Community extends what individuals can achieve alone. Tool adoption requires counterparties, and trust develops through repeated interaction. Local meetups function as physical bridgeheads where online handles become remembered faces and the parallel economy grows through recurring trade. Implementation moves beyond the solitary user as well. Households and small businesses need shared rules: which communication channels to use for sensitive matters and how treasury keys separate from daily operational

access. The first meaningful win is a commercial loop that earns and spends privately, then repeats.

Proxy merchants help that loop start sooner. OTC exchangers and purchasing agents let small networks trade before every participant can stand entirely on private rails.

No single tool or configuration provides adequate protection on its own. Tools serve threat models, and generic advice assumes generic threats, effective protection requires the personalized assessment Chapter 22 develops. The parallel economy is also not ready for complete exit from state-supervised systems. Implementation is incremental, and today's parallel economy supplements the old system instead of replacing it. Chapter 25 synthesizes the full argument, assesses what the working tools have already achieved and what remains unbuilt, and answers the book's opening challenge directly. Privacy is an ongoing practice; the question to ask is whether you are making progress from where you are now, and the answer depends on starting there, improving what can be improved, and finding the community that makes the practice sustainable.

Notes

[1]Moxie Marlinspike, "The Network in Motion," Moxie Marlinspike (blog), 2016, `https://moxie.org/2016/01/17/network-protocols.html`. Marlinspike co-founded Signal and writes on cryptographic protocol design; the essay argues that static protocol assumptions break as the network layer evolves under them.

[2]For threat modeling methodology accessible to non-specialists, see "Your Security Plan," Electronic Frontier Foundation, Surveillance Self-Defense guide, `https://ssd.eff.org/en/module/your-security-plan`. For the formal security-engineering treatment, Adam Shostack, *Threat Modeling: Designing for Security* (Wiley, 2014), covers STRIDE and related frameworks and is the standard reference. Ross Anderson, *Security Engineering*, 3rd ed. (Wiley, 2020), cited at Chapter 22, note 2. For a shorter practitioner guide oriented to journalists and activists, see the Freedom of the Press Foundation's digital security training materials at `https://freedom.press/training/`.

[3]Micah Lee, "A DIY Guide to Feminist Cybersecurity," *The Intercept*, November 12, 2017, `https://theintercept.com/2017/11/12/a-diy-guide-to-feminist-cybersecurity/`, is the best short introduction to progressive privacy implementation for readers without a technical background. Martin Shelton maintains current step-by-step recommendations at

https://martinshelton.medium.com. The Electronic Frontier Foundation's Surveillance Self-Defense guide at https://ssd.eff.org bundles threat-model walkthroughs, tool tutorials, and tactical scenarios ("Attending a Protest," "Crossing the US Border," etc.). For the most aggressive tactical separation, Michael Bazzell, *Extreme Privacy: What It Takes to Disappear*, 5th ed. (2023), is the definitive manual; Bazzell's accompanying podcast, *The Privacy, Security, and OSINT Show*, keeps the techniques current.

[4]KeePassXC (https://keepassxc.org) is a cross-platform open-source password manager storing a local encrypted database; Bitwarden (https://bitwarden.com) is a cloud-syncing open-source alternative with optional self-hosted servers (Vaultwarden, https://github.com/dani-garcia/vaultwarden, is the popular Rust-based compatible self-hosted server). For methodology and threat-model walkthroughs, see the Electronic Frontier Foundation's Surveillance Self-Defense guide on password managers at https://ssd.eff.org/en/module/animated-overview-using-password-managers-stay-safe-online. For password-based authentication failure modes at scale, see Troy Hunt's work at https://haveibeenpwned.com. For passkeys and the WebAuthn/FIDO2 standard that backs hardware-bound authentication, see the W3C specification at https://www.w3.org/TR/webauthn-3/ and the FIDO Alliance overview at https://fidoalliance.org/how-fido-works/.

[5]SIM-swap attacks occur when an adversary socially engineers or bribes a mobile carrier employee to transfer a victim's phone number to a SIM card the adversary controls, redirecting all SMS messages including authentication codes. The attack bypasses SMS-based two-factor authentication entirely. High-profile SIM-swap cases include the 2019 Twitter account hijackings and the 2022 Coinbase breach enabling large-scale account takeovers. For a detailed technical and social analysis, see Rachel Tobac and the SocialProof Security team's documentation; also Joseph Cox and Lorenzo Franceschi-Bicchierai, "T-Mobile Hack," *Vice Motherboard*, 2021; and the FTC's guidance at https://www.ftc.gov/business-guidance/resources/ftcs-guide-protecting-consumers-sim-swap-scams. The FIDO2/WebAuthn hardware-key standard was designed specifically to be structurally immune to this class of attack because the credential is bound to a domain by cryptographic hardware, not delivered over a telephony channel.

[6]Foundation-level software referenced in this paragraph: Signal (https://signal.org) is the end-to-end encrypted messenger treated in Chapter 17; its Double Ratchet protocol specification is at https://signal.org/docs/specifications/doubleratchet/. ProtonMail (https://proton.me/mail) is a Swiss-based email provider with client-side encryption for messages between Proton users and with OpenPGP support for messaging external addresses. Tutanota (rebranded as Tuta, https://tuta.com) is a German-based email provider with end-to-end encryption for messages between Tuta users. For at-rest disk encryption, FileVault is Apple macOS's built-in full-disk encryption, BitLocker is Microsoft Windows's,

and LUKS (Linux Unified Key Setup, https://gitlab.com/cryptsetup/cryptsetup) is the Linux disk-encryption standard used by most major distributions. Firefox (https://www.mozilla.org/firefox/) with a privacy-hardened user.js configuration (arkenfox-user.js at https://github.com/arkenfox/user.js is the most-maintained profile) reduces fingerprinting and telemetry surface; Brave (https://brave.com) is a Chromium fork with built-in tracker and fingerprint blocking. uBlock Origin (https://github.com/gorhill/uBlock) is the reference open-source content-blocker. For content encryption tooling more generally, GnuPG (https://gnupg.org) remains the OpenPGP reference implementation, and age (https://age-encryption.org), designed by Filippo Valsorda, is the modern small-surface-area file encryption tool with a much simpler threat model than PGP.

[7]Mullvad Browser is Tor Browser's anti-fingerprinting work packaged for use over any network; LibreWolf is Firefox with arkenfox defaults pre-applied. Both prioritize rendering the browser fingerprint indistinguishable from other users of the same browser. Mullvad Browser, https://mullvad.net/en/browser, a collaboration between the Tor Project and Mullvad. LibreWolf, https://librewolf.net/, with source at https://codeberg.org/librewolf. Arkenfox user.js profile, https://github.com/arkenfox/user.js. For the fingerprinting threat model these browsers address, see Peter Eckersley, "How Unique Is Your Web Browser?" *Privacy Enhancing Technologies Symposium 2010*, and EFF Cover Your Tracks, https://coveryourtracks.eff.org/. Tor Browser, https://www.torproject.org/download/, remains the reference standard for anonymity-over-Tor browsing and ships the same fingerprinting defenses Mullvad Browser inherits.

[8]SimpleLogin (Proton AG) and Addy.io are open-source self-hostable email-aliasing services; DuckDuckGo Email Protection and Apple Hide My Email are convenience plays where the forwarding provider sees plaintext email; the structural benefit is a different forwarding address per service so a breach at any one service cannot be cross-correlated. SimpleLogin, https://simplelogin.io/, with source at https://github.com/simple-login/app. Addy.io (formerly AnonAddy), https://addy.io/, with source at https://github.com/anonaddy/anonaddy. DuckDuckGo Email Protection, https://duckduckgo.com/email/. Proton Pass (with SimpleLogin integration), https://proton.me/pass. Apple Hide My Email, https://support.apple.com/en-us/HT210425. Mozilla's Firefox Relay, https://relay.firefox.com/, is the remaining mainstream alternative. For the underlying universal-tracker critique that motivates email aliasing, see Arvind Narayanan and colleagues' OpenWPM work at Princeton, https://webtransparency.cs.princeton.edu, documenting cross-service email-based tracking.

[9]The Bitcoin Wiki's privacy page at https://en.bitcoin.it/wiki/Privacy is the best-maintained encyclopedic reference for Bitcoin-specific privacy practice, covering address reuse, change detection, common-input-ownership heuristics, CoinJoin variants, and Lightning-layer privacy. For practitioner-oriented treatments, 0xB10C's ongoing analysis at https://b10c.me/ and https://transactionfee.info covers chain-analysis techniques and defenses. Sparrow Wallet's documentation at https://sparrowwallet.com/docs covers coin control, label hygiene, and PayJoin integration. Chapter 20 develops the full transaction-graph threat model; note 14 of that chapter lists the full set of working tools.

[10]Peer-to-peer Bitcoin exchanges allow buyers and sellers to trade directly without a centralized intermediary holding funds or collecting identity documents. Active platforms include Bisq (https://bisq.network/), a decentralized application using multisig escrow and dispute arbitration with no central server; RoboSats (https://learn.robosats.com/), a Lightning-native P2P exchange using hold invoices and a coordinator for dispute resolution; Peach Bitcoin (https://peachbitcoin.com/), a mobile-first European P2P platform; and AgoraDesk/LocalMonero (now closed) for the Monero equivalent. Bitcoin ATMs that operate without KYC requirements are catalogued at https://coinatmradar.com/ with filter options for verification level, though regulatory pressure has pushed most US ATMs above the KYC threshold. For the surveillance-infrastructure critique of KYC exchanges, see Chapter 10 and the Financial Action Task Force (FATF) guidance on virtual asset service providers at https://www.fatf-gafi.org/en/publications/Fatfgeneral/Virtual-assets-guidance.html, which is the international standard driving KYC requirements.

[11]Zeus is a self-custodial Lightning wallet that connects to a user-operated Lightning node, preserving payment privacy by keeping routing information on the user's own node rather than a provider's infrastructure; https://zeusln.app/, with source at https://github.com/ZeusLN/zeus. Phoenix is a self-custodial Lightning wallet using an LSP (ACINQ) for channel management that simplifies setup at the cost of routing visibility to the provider; https://phoenix.acinq.co/, with source at https://github.com/ACINQ/phoenix. For the broader Lightning privacy model and LSP trust assumptions, see lightningprivacy.com and Chapter 20, note 12.

[12]Mullvad VPN, https://mullvad.net/, is a Swedish no-log VPN provider accepting anonymous payment; IVPN, https://www.ivpn.net/, is a Gibraltar-registered no-log provider with a published warrant canary. Both are among the providers most frequently recommended in independent audits. For independent VPN audit references, see Mullvad's published audits at https://mullvad.net/en/blog/tag/audits and IVPN's at https://www.ivpn.net/blog/. On the structural limitation that all VPNs

shift rather than eliminate the trust relationship, see the EFF's explainer at https://ssd.eff.org/en/module/choosing-vpn-thats-right-you.

[13]Nym is a mixnet protocol providing stronger metadata protection than Tor's low-latency onion routing by batching and reordering packets across mix nodes; https://nymtech.net/, with protocol specification at https://github.com/nymtech/nym. Tor Browser, the reference tool for web anonymity over the Tor network, is at https://www.torproject.org/download/. Tails is an amnesic live operating system that boots from USB, routes all traffic through Tor, and leaves no persistent trace on the host; https://tails.boum.org/. For the anonymity-set and timing-attack differences between Tor and mixnet designs, see Claudia Diaz, "Measuring Anonymity," *Privacy Enhancing Technologies Symposium 2002*.

[14]Whonix is a desktop operating system designed for anonymity, running as two virtual machines: a Gateway VM that routes all traffic through Tor, and a Workstation VM that is entirely isolated from the network except through the Gateway. Even if the Workstation is compromised, the attacker cannot learn the user's real IP address because the Workstation has no direct network access. Whonix runs on top of a host OS (typically Debian or Qubes OS) and is available as KVM/VirtualBox appliances. Project site: https://www.whonix.org/, with source at https://github.com/Whonix. When Whonix runs inside Qubes OS (Qubes-Whonix), the two approaches combine: Xen-based compartmentalization from Qubes and Tor-based network anonymization from Whonix. For the threat model and design rationale, see https://www.whonix.org/wiki/Main_Page and Patrick Schleizer's technical documentation at https://www.whonix.org/wiki/Dev/Design-Principles.

[15]GrapheneOS features documentation at https://grapheneos.org/features. GrapheneOS documents sandboxed Google Play, its hardened_malloc allocator, and hardware memory tagging on supported devices as distinct features. Current GrapheneOS support is limited to Google Pixel devices, which provide the verified-boot and security-hardware support the project targets. Qubes OS, https://www.qubes-os.org/, isolates applications in separate Xen-based virtual machines; hardware compatibility requirements (VT-x/VT-d) and the certified hardware list are at https://www.qubes-os.org/hcl/. See also Synacktiv, "Exploring GrapheneOS Secure Allocator: Hardened Malloc" (2024).

[16]Mail-in-a-Box, https://mailinabox.email/, is a one-click self-hosted email server bundling Postfix, Dovecot, and spam filtering; source at https://github.com/mail-in-a-box/mailinabox. Nextcloud, https://nextcloud.com/, is the leading self-hosted file sync, calendar, and collaboration platform; source at https://github.com/nextcloud/server. Both require ongoing maintenance: security patches, backup discipline, and DNS configuration. For a broader survey of self-hostable privacy-respecting

alternatives to cloud services, see https://github.com/awesome-selfhosted/awesome-selfhosted.

[17]I2P (Invisible Internet Project) is a peer-to-peer anonymizing overlay network providing unidirectional tunnels, garlic routing (bundling multiple messages together), and a distributed network database; it is optimized for internal services (eepsites, I2P-native applications) rather than clearnet access and is therefore complementary to Tor rather than a replacement. Project site: https://geti2p.net/, with source at https://github.com/i2p/i2p.i2p. Java I2P is the reference implementation; i2pd (https://i2pd.website/, https://github.com/PurpleI2P/i2pd) is the C++ implementation with lower resource requirements. For a technical comparison of Tor and I2P threat models, see the EFF's resources and the I2P project's own threat-model documentation at https://geti2p.net/en/docs/how/threat-model. The sovereign-host configuration in section 23.3 can expose services over I2P hidden services as an alternative to Tor onion services.

[18]Start9 and Umbrel are the two leading personal-server distributions packaging self-hosted applications (Bitcoin node, Nostr relay, Matrix server, Nextcloud, etc.) with privacy-preferring defaults; the reference hardware is a Raspberry Pi 5 or Intel NUC-class miniature PC, but a repurposed old laptop with a spare SSD runs the same workloads equally well, and a Bitcoin full node runs on the same box so the wallet queries the user's own node instead of a commercial provider. Start9, https://start9.com/, with source at https://github.com/Start9Labs. Umbrel, https://umbrel.com/, with source at https://github.com/getumbrel. Yunohost, https://yunohost.org/, is the third major option in the category, with source at https://github.com/YunoHost. For Bitcoin full-node operation, Bitcoin Core at https://bitcoincore.org/ and curated guides at https://github.com/jlopp/civil-disobedience (Jameson Lopp's bitcoin-self-hosting resources). For Nostr relays, strfry (https://github.com/hoytech/strfry) is the most-deployed implementation. For threat-model boundaries on residential-IP hosting, Matt Odell and Jameson Lopp's ongoing commentary at https://citadeldispatch.com/ and https://www.lopp.net/ covers the practical tradeoffs.

[19]Local inference on open-weight LLMs runs the computation on the user's own hardware with no prompt leaving the device; llama.cpp and Ollama are the reference developer stacks, LM Studio and GPT4All the consumer frontends, Apple MLX and Foundation Models the on-device stacks on Apple Silicon. The category has matured in the last several years without yet matching frontier hosted capability. llama.cpp, Georgi Gerganov, https://github.com/ggerganov/llama.cpp, with the GGUF format specification at https://github.com/ggerganov/ggml/blob/master/docs/gguf.md. Ollama, https://ollama.com/, with source at

https://github.com/ollama/ollama. LM Studio, https://lmstudio.ai/. GPT4All, Nomic AI, https://gpt4all.io/. Jan, https://jan.ai/. Apple MLX framework, https://github.com/ml-explore/mlx, and Foundation Models API documentation at https://developer.apple.com/documentation/foundationmodels/. Nvidia NIM, https://www.nvidia.com/en-us/ai/. For the privacy-preserving machine-learning research broadly, Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov, "Membership Inference Attacks Against Machine Learning Models," *IEEE Symposium on Security and Privacy 2017*, is the foundational paper on the attack class that local inference sidesteps entirely.

[20]For readers who want a step-by-step applied companion to this chapter, Michael Bazzell's *Extreme Privacy: What It Takes to Disappear*, 5th ed. (2023) is the definitive tactical reference on disentangling identity from surveillance infrastructure. For the broader context, Carissa Véliz, *Privacy Is Power* (2020) and Shoshana Zuboff, *The Age of Surveillance Capitalism* (2019) argue the political stakes; Bruce Schneier, *Data and Goliath* (2015) surveys the infrastructure. On opting out of specific surveillance classes: Kashmir Hill's reporting at *The New York Times* is consistently excellent on data-broker ecosystems, and the EFF's Surveillance Self-Defense guide (ssd.eff.org) stays current on the tooling side.

[21]On the strategy literature that treats meetups, merchant rooms, OTC exchange, proxy merchants, and security culture as the bridge between individual privacy practice and a functioning parallel market, see Smuggler and XYZ, *The Second Realm: Book on Strategy* (Liberty Under Attack Publications, 2018); Kyle Rearden and Shane Radliff, *Just Below the Surface: A Guide to Security Culture* (Independently Published, 2019); and Shane Radliff and Tom Marshall, *Going Mobile* (*#VanLife Zine*) (CreateSpace Independent Publishing Platform, 2019). These are best read as practical and strategic companions to this chapter, not as replacements for the primary technical references cited above.

[22]Nostr (Notes and Other Stuff Transmitted by Relays) is a decentralized, censorship-resistant social protocol in which users control their identity through cryptographic keypairs; https://nostr.com/, with protocol specification at https://github.com/nostr-protocol/nostr. The protocol is treated in depth in Chapter 21. Matrix is a federated messaging protocol with end-to-end encryption; the reference client is Element (https://element.io/), and the specification is at https://matrix.org/.

[23]Matrix is a federated, open-standard messaging protocol supporting end-to-end encryption via the Olm/Megolm double-ratchet implementation; the specification is at https://matrix.org/. Element (https://element.io/) is the reference client maintained by Element Matrix Services. The Matrix protocol allows self-hosted homeservers that federate with the broader network, making it harder to shut down than a centralized service. SimpleX Chat (https://simplex.chat/, source at https://github.com/simplex-chat/simplex-chat) takes a different architectural approach: no user identifiers of any kind, no phone number or email

required, and one-time invitation links as the contact mechanism; messages are routed through user-chosen relay servers. Briar (`https://briarproject.org/`, source at `https://code.briarproject.org/briar/briar`) is a peer-to-peer encrypted messenger that routes traffic over Tor and can synchronize over Bluetooth or WiFi when no internet connection is available, making it suitable for high-censorship or infrastructure-disrupted environments. These three tools represent different points in the tradeoff space between usability, federation, and server-trust minimization.

[24]Robert Axelrod, *The Evolution of Cooperation* (Basic Books, 1984), remains the classic game-theoretic treatment of how cooperation stabilizes among self-interested actors through repeated interaction. Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990), empirically documents community trust mechanisms in real-world commons governance and is the canonical source for non-state collective-action institutions. On reputation in online and pseudonymous contexts, Paul Resnick and Richard Zeckhauser, "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System," in *The Economics of the Internet and E-Commerce*, ed. Michael Baye (Elsevier, 2002), is the foundational empirical paper. Chrysanthos Dellarocas, "The Digitization of Word of Mouth," *Management Science* 49, no. 10 (2003): 1407–1424, surveys reputation-mechanism design. Chapter 24 develops pseudonymous reputation, credentials, and decentralized institutions in depth.

[25]Angor is a working implementation of the multisig-plus-staged-disbursement pattern at the investment-coordination scale. Each project is funded under a 2-of-2 multisignature contract between investor and founder, with disbursement gated on time-locked milestones rather than released up front; the founder shows progress at each stage and the investor retains the option to halt remaining tranches. Project metadata and the investor-founder communication channel run on Nostr instead of on a hosted intermediary, and the on-chain side uses Bitcoin with Taproot to keep the contract structure private to the participants. The protocol is open and documented at `https://angor.io` with the technical specification at `https://docs.angor.io` and source at `https://github.com/block-core/angor`; the application is at `https://hub.angor.io`. Angor extends the escrow logic this section describes from single trades into multi-stage commercial relationships, and is the closest currently deployed analogue to staged-trust commerce among pseudonymous counterparties.

[26]ShopinBit operates the proxy-merchant pattern at retail and concierge scale from Germany, in operation since December 2018, accepting Bitcoin and Monero in exchange for goods that ShopinBit procures from conventional fiat-priced suppliers and ships to the customer. The privacy work the model does is identity-shielding at the supplier interface: ShopinBit holds the KYC relationship with the upstream vendor, and the end customer never has to expose identity to that vendor. Customer accounts are optional, no marketing analytics are loaded

by the storefront, customer data is deleted automatically after thirty days, and crypto payment carries a three percent discount over fiat. The Concierge service extends the same logic to anything sourceable on the open market, including cars, travel bookings, and B2B procurement, with the request submitted through a chat interface that requires no email or account. The proxy-merchant pattern handles a different problem from the escrow mechanics this section describes: the supplier requires identity the customer does not wish to provide, so the proxy bridges the gap by accepting crypto from the customer and presenting conventional payment and identity upstream. Other examples of the category include Bitrefill, which sells gift cards redeemable at conventional retailers, and a range of regional Bitcoin-acceptance retailers operating on similar lines. See `https://shopinbit.com` and the Concierge documentation at `https://shopinbit.com/Concierge/`.

[27]OTC (over-the-counter) Bitcoin exchangers trade directly with counterparties rather than through an order book, typically accepting cash or bank transfer in exchange for Bitcoin without mandatory identity verification. The category ranges from informal local traders found through community meetups to semi-professional services advertising on platforms such as Bisq (`https://bisq.network/`) and Peach Bitcoin (`https://peachbitcoin.com/`). For the proxy-merchant model as a privacy-preserving intermediary layer, see Smuggler and XYZ, *The Second Realm: Book on Strategy* (Liberty Under Attack Publications, 2018), which develops the logic of specialization-at-the-surveillance-boundary. Bitrefill (`https://www.bitrefill.com/`) is the largest retail proxy-merchant in the category, selling gift cards and top-ups redeemable at conventional fiat-priced merchants in exchange for Bitcoin or Lightning payment, enabling Bitcoin holders to purchase from retailers that do not accept Bitcoin directly without disclosing their identity to those retailers.

24

TRUST AND DISPUTE IN THE PARALLEL ECONOMY

> "Reputations will be of central importance, far more important in dealings than even the credit ratings of today."
>
> Timothy C. May, *The Cyphernomicon* (1994)[1]

The preceding chapter showed individuals and small groups building privacy practice from the ground up: threat modeling, tool selection, community formation, and the first commercial loops within trusted circles. That is where adoption begins. The question Chapter 23 leaves open is what happens when the circle widens. How do strangers trade, and how do they resolve uncertainty, when they refuse the state's identity system and do not trust the state's courts?

Privacy is a condition of action, and surveillance distorts exchange and weakens calculation. Cryptographic systems can raise the cost of control until many attacks stop paying; Bitcoin gives money a resistant base; anonymous networks protect communication; zero-knowledge proofs let people prove narrow facts without opening the whole file; and Nostr gives users identities they control on a social medium no platform owner can seize.

Answering that question requires more than tools. The cryptographic stack moves bits and settles payments, but it does not establish reliability across time or adjudicate what happened when a

deal fails. Those are institutional functions, and a parallel economy needs institutions that perform them without routing through the state's identity system or its courts.[2]

24.1 Why Tools Alone Do Not Make a Market

Some exchanges need almost no trust. When you buy tomatoes from a farmer and hand over cash at the same moment the tomatoes pass into your hands, the transaction closes on the spot: payment and delivery happen together, and each side walks away with final settlement.

The trust problem enters as soon as settlement stretches through time or distance. A merchant asks for payment now and promises shipment tomorrow. A developer sells updates that will arrive over the next year, while a founder wants staged funding before the product is finished. A pseudonymous seller may also claim a record of honest dealing that the buyer cannot yet verify.

Encryption does not answer those questions, and neither do Bitcoin or Nostr by themselves. These tools solve transport, identity control, publication, and settlement. Reliability across time and the distinction between fraud and honest error belong to other mechanisms.

The modern state bundles several distinct functions into one package and then presents the package as indispensable. It ties identity to legal documents and channels disputes into state courts, all backed by regulators and police. It also records firms and preserves a final enforcement backstop. Markets have long relied on some of these functions. They do not need them to remain under state monopoly.

A parallel economy does not need to abolish trust. It needs ways to create trust without surrendering everyone to permanent identification and surveillance. It needs ways to narrow trust, price trust, distribute trust, and contain the damage when trust fails.

That is an institutional problem. Timothy May saw this early. Anonymous business and electronic contracts were possible, he argued, but reputations would become central. Eric Hughes made the

same point from another angle: privacy in an open society requires anonymous transaction systems, and each party to a transaction should know only what is directly necessary for that transaction.[3] Nick Szabo pushed the analysis farther. In a world of digital commerce, he wrote, the dominant costs move toward jurisdiction, trust, and security. Search, negotiation, commitment, performance, and adjudication all need structure.[4]

The technical stack in this book now exists. What follows is the question May and Szabo left hanging in the air: how does that stack support commerce between people who know each other only through keys and a record of performance?

24.2 Reputation Under Pseudonymity

Reputation needs continuity. A buyer must be able to say: I traded with this seller before, the goods arrived, the quality matched the claim, the messages were prompt, the refund came when it should have come. Memory requires continuity, and reputation requires memory.

Strong anonymity cuts against that continuity. If each interaction uses a fresh identity with no link to any prior history, observers learn less, but counterparties learn less too. The same unlinkability that blocks surveillance also blocks accumulated trust.

This tension does not disappear because we dislike it. A system that promises strong anonymity and durable reputation in the same channel is usually hiding one of the tradeoffs. The working compromise in many markets is pseudonymity.

A persistent pseudonym can accumulate history while still refusing civil identity. What the key has to do is persist long enough for other people to observe conduct across time; what it does not have to do is map to a passport or a state registry. The pseudonym becomes expensive to abandon because abandoning it means burning accumulated trust. That cost disciplines behavior. A seller who has spent two years building a name on Nostr, answering buyers, shipping products, refunding mistakes, and posting under the same

key has something to lose if he cheats. His identity is economic, not a legal fact. That is enough.

Nostr matters here because it makes ordinary social behavior legible as signed evidence. Follows, replies, reposts, mutes, zaps, long-form posts, software releases, and marketplace listings can all attach to the same key. None of these signals is perfect, and all can be gamed at the margin. The overall pattern still carries information.

An account followed by people you already trust is not proven honest, but it is easier to take seriously. A merchant with years of public history under one key is not guaranteed to perform, but the cost of defection is higher than for a fresh key with no record. A developer whose releases are signed by the same identity that publishes essays, receives criticism, answers bugs, and gets endorsed by other developers offers a thicker signal than a nameless download link.

The signal can also stay local. Reputation does not need to be one global score handed down by a central authority; trust is contextual, and you may trust one account as a developer and a different one as a merchant. My graph differs from yours, and that difference is a feature, not a weakness. Centralized platforms flatten trust into one badge, while market systems let trust remain distributed and plural.

There is still a cold-start problem. New entrants begin with no history, so they typically start with small trades, introductions, or domain-linked proof of who they are in some narrower context. Some lean on escrow until they build a name; others earn trust inside a local community before trying to sell to strangers. None of this is frictionless, and commerce with no gatekeepers still has filters. The filters just move from permission granted by administrators to judgment exercised by participants themselves.[5]

Negative history creates a second pressure. A seller who burns one pseudonym can open another and return as a newcomer, which leaves purely local reputation systems weak against repeat fraud. Parallel markets need enough continuity to make defection costly without turning every market identity into a universal identifier.

Zero-knowledge proofs open a partial answer to this tension, though practical deployment is still early. Persistent pseudonyms

resolve the tension between anonymity and continuity by keeping the same key alive, which means everything under that key becomes linkable in exchange for a visible track record. A different compromise becomes possible when reputational claims can be separated from the specific transactions and identities that produced them. Attestation-based protocols where attesters sign statements about a user that the user later proves under zero knowledge make it conceivable to claim "I have completed at least N trades at rating at least R, with people whom you follow." or "I have been vouched for by K members of this trader cohort" without revealing which trades or which other identities are yours.[6] These systems still depend on attesters who observe the underlying facts, so cold-start and sybil resistance do not disappear; collusion among enough attesters can still reassemble a trail. As the tooling matures, the default shape of reputation could shift: claims accumulated under pseudonyms and presented at the point of use, with each interaction revealing only the piece the transaction needs.

That movement is healthy. The state wants reputation tied to civil identity because civil identity is easy to tax and easy to coerce. A parallel economy wants reputation tied to performance. Performance is the information that trade needs.[7]

24.3 Credentials and Selective Disclosure

Reputation is soft evidence. It grows from observed behavior. Many transactions need something harder.

One buyer may need proof that the seller controls a known domain. Another user may need proof that a release came from the developer who claims authorship. In other cases the relevant fact is continuity of the same key across two channels, or proof that a newcomer was vouched for by an existing member and meets a rule that governs participation.

The state solves these problems through over-disclosure. It demands the full packet: passport, driver's license, utility bill, customer profile. The institution gets a file thick enough to do far more than the immediate transaction requires.

That is bad design.

Most transactions need a narrow fact. The verifier does not need the whole person. The verifier needs only the piece of information on which the transaction turns.

Nostr already supplies one simple model. NIP-05 lets a pubkey be associated with a DNS- and HTTPS-backed identifier without asking a platform for a badge.[8] The trust still lies in the domain and in the entity behind it, and in the web PKI serving the `nostr.json` file. A company can attest that a given key belongs to one of its developers. An individual can attest that a given key is the one tied to his own website. The system does not abolish trust. It narrows the trust to the specific claim.

The same logic extends farther. A known key can sign a one-time challenge and send the response through another channel, so that the verifier learns the person on the anonymous channel controls the same key whose public history he already knows. The verifier does not learn a legal name and does not need one, because the fact that matters for this interaction is continuity between a known reputational identity and the present conversation.[9]

Zero-knowledge systems push the same principle into richer credential problems. Chapter 15 already showed the general pattern. A system can prove age above a threshold without revealing birthdate. It can also prove membership or control of a credential without handing over the credential itself. For the parallel economy, this replaces bulk identity extraction with narrow proof.

The goal is selective disclosure: reveal control of the developer key, not a home address; reveal eligibility under the age rule, not a birthdate. Standing in a market community follows the same discipline: show the claim the transaction requires, not the web of relationships behind it.

Selective disclosure does not remove all risk. Credentials can become surveillance infrastructure if they are logged or centralized, especially when they become compulsory across contexts. A pseudonymous market can quietly rebuild the same dossier system it claimed to reject if every trade demands one more signed proof

and each proof gets stored forever. The answer is still discipline. Reveal what the transaction needs and no more.

That rule links the cypherpunk side of the book to the praxeological side. The transaction determines the information requirement. Anything beyond that requirement is surplus extraction.

The tooling supporting this principle has reached production over the last several years. Group-membership attestations through protocols such as Semaphore have been deployed at conference-scale events for ticketing and credential presentation, with the cryptographic primitives carrying hundreds to thousands of users per deployment. Wallet-creation flows that derive a pseudonymous wallet from an existing identity-provider account, without the provider learning the derivation, are in production on at least one blockchain. Verifiable-credential standards through the World Wide Web Consortium provide a generic format in which selective-disclosure proofs can be expressed. The state-issued side of the same infrastructure is also maturing, with mobile driver's licenses based on an international standard now accepted at identity checkpoints in several countries.[10]

The book's framework insists on a qualification that the standards documents do not. A credential system that ships with selective disclosure as a feature is not the same as a credential system that is used with selective disclosure as a practice. The state-issued variants tend to use selective disclosure as a convenience (faster verification at an airport) instead of as a privacy primitive (minimal exposure in routine life), and the verifier-side incentives push toward requesting more than the transaction requires. The pseudonymous-community variants tend to use selective disclosure as intended, because the participants care about the underlying principle. The tooling is increasingly the same; the posture of the users is what determines whether the tooling protects privacy or rebuilds the dossier with extra steps.

24.4 Escrow and Simple Dispute Resolution

Some transactions do not need a court. They need a release condition.

If buyer and seller are strangers and the goods will ship later, one simple answer is escrow. Funds sit under the control of a third party until delivery is confirmed, or under a multisignature arrangement that requires agreement between the trading parties and an arbiter. The buyer no longer needs blind faith, and the seller no longer ships with no assurance of payment, so the trust surface shrinks on both sides.

Bitcoin makes these arrangements concrete. A two-of-three multisignature setup can split keys between buyer and seller, with a third key held by an arbiter. If the trade finishes smoothly, buyer and seller release the funds together. If trouble arises, the arbiter joins one side and the funds move with two signatures. No one person has unilateral control. That is already a better structure than the ordinary platform model where a company takes the money and writes the rules.[11]

Small markets often layer social practices on top of the cryptography. New counterparties trade in small size first, and sellers ship one item before asking for larger orders. Communities recommend known escrow agents whose own names are worth protecting. Repeat trade lowers uncertainty because each successful exchange becomes evidence for the next, and the accumulated evidence is itself part of what new entrants are later offered as social proof.

These mechanisms work because many disputes are simple. Sometimes the question is whether a package arrived, whether a repository shipped the promised release, whether funds cleared, or whether one side violated an agreed rule visible to both parties. In such cases, clear release conditions do much of the work.

The gain is real but should be stated with precision. Escrow and multisig do not eliminate trust; they reallocate and narrow it. The parties still trust the keys and the software, and some arrangements also require trust in an arbiter. Yet narrowed trust is often enough, because many market problems do not require perfect assurance but

only a structure that makes fraud costly and honest trade easier than cheating.

Staged commitments begin to matter here as well. A funding arrangement that releases money in stages gives investors bargaining power throughout the life of the project. A buyer who can recover funds not yet released has a real position from which to negotiate. A service contract that encodes partial payment and conditional release brings discipline into the contract itself. Here the contract starts to move out of paper and into protocol. The machine does not replace judgment. It holds judgment inside a tighter frame.

24.5 Arbitration and Contested Facts

Escrow works well when the disputed fact is clean. It works less well when the conflict turns on quality disputes or evidence no script can settle.

A package may show delivered even though the buyer says the box was empty. A contractor may insist the job met the agreed spec while the client says the work is unusable. In another case the seller defends the product description or the developer claims a milestone was met, while the buyer or funders say the delivered result does not match the promise.

These are judgment problems sitting on top of the settlement layer, and settlement alone cannot solve them.

Modern courts bundle several functions at once: they collect evidence, define procedure, interpret contested facts, produce judgments, and connect those judgments to force. Much of the value in the bundle does not come from the monopoly itself. It comes from the existence of procedures, records, standing, and a forum that parties expect other parties to recognize.

A parallel economy must unbundle this stack, and some parts are already available. Reputation can punish obvious fraud and escrow can freeze funds until a ruling arrives; community standing can pressure a dishonest merchant or a corrupt arbiter into better behavior. In bounded communities, exclusion can be a severe sanc-

tion; a merchant who cheats may keep one payment and lose the next hundred, which is often enough to keep ordinary trade honest.

Harder cases remain underbuilt.

A serious arbitration system needs more than a chat room and a wallet.[12] It needs procedures for submitting evidence and rules for weighing testimony. It also needs standards for conflicts of interest, along with some way to make precedent legible so each new dispute does not begin from zero. Participants also need reason to believe that arbiters have something to lose if they rule badly. In some contexts a posted bond may do that work. In others, repeated reputation or technical standing will carry more weight.

Merchants have long relied on private ordering and repeated dealing, often alongside municipal or state courts, to discipline trade where formal law was slow or poorly matched to the commerce at hand. Digital networks change the scale and the tooling without changing the underlying problem, which is that trade among strangers still needs a way to handle breach. One contemporary response is to codify an explicit libertarian private-law baseline that voluntary adopters can reference in their contracts and arbitration clauses, supplying the shared legal starting point that the unbundled layer still lacks.[13]

This layer remains immature.

There are experiments with bonded arbiters and community juries, along with a range of protocol-governed mechanisms.[14] Some work for narrow cases. None yet carries the legitimacy or procedural maturity of a court system backed by long practice. That gap should not be hidden. The parallel economy has solved settlement more cleanly than it has solved adjudication.

Yet the direction is clear. Courts are institutions, and institutions can compete, specialize, fail, improve, and split into narrower services. The parallel economy has not built a full private-law order. The relevant question is whether the pieces needed for one can emerge from repeated exchange under reputational pressure and cryptographic control of funds. They can.

24.6 Merchant Discovery and Institutional Emergence

Markets do not appear because someone declares them open. Markets appear when people find each other and trade again.

That is why the institutional layer is larger than escrow and arbitration. It includes discovery.

A merchant with a good product is useless to the buyer who cannot find him. The same holds for a skilled developer whose release cannot be distinguished from a counterfeit, or for a fair arbiter unknown to the parties who need him. A market grows when the paths between these roles become cheap and legible.

Nostr helps here because one identity can carry across several contexts at once. The same key that posts notes can sign software releases, publish marketplace listings, receive payments, answer questions, and show a long public history of conduct. A buyer does not inspect one isolated listing and decide in a vacuum. The buyer can inspect the whole trail: who follows this seller and how long the key has been publishing, and whether complaints were handled in public view or answered with silence.

The result is thickness. A thick market preserves information about conduct across time and raises the cost of losing standing as repeated exchange grows denser. It also lets counterparties compare alternatives. A thin market sends the participants back to zero after each failure.

That thickening happens through repeated contact. Meetups, workshops, merchant circles, private group chats, bug reports, escrow recommendations, and signed listings all contribute. Digital and physical institutions reinforce each other. A merchant you met in person is easier to trust online. A developer whose code you used for two years is easier to trust when he begins selling hardware. A local Bitcoin meetup that turns into recurring trade stops being a social event and becomes infrastructure.

Proxy Merchants and Bridge Roles

Markets also thicken by specialization. One of the most useful roles at the edge of a parallel economy is the proxy merchant: a

person or small firm that interfaces with the exposed side of the economy so other participants can stay off that boundary. Sometimes the role is an OTC exchanger moving between bitcoin and local currency. Sometimes it is a purchasing agent who buys from legacy-facing suppliers on behalf of a counterparty who wants to keep his name and address off the order, or the operator of pickup points and forwarding routes that keep buyer and seller from meeting directly. The function stays the same. The arrangement concentrates exposure in a chosen specialist who prices the risk and builds a reputation for handling it well.[15]

These roles narrow trust and put a price on it. They also create a familiar danger, because a bridge used by everyone becomes a bridge watched by everyone. The institutional answer is the same one the chapter has already given for escrow and arbitration: plurality and easy replacement. A healthy parallel market keeps many proxy merchants and related bridge services in play so that no single operator becomes a bottleneck.

The phrase parallel economy means what it says: a second set of rails on which communication, payment, reputation, discovery, and trade can run. These rails may exist inside the same city and on the same streets as the old system, and they remain distinct when trade and information flows no longer depend on state-controlled intermediaries.

Institutional emergence follows use. No committee can draft the full order in advance. Merchants with good dispute practices get more business, and escrow providers who handle conflicts fairly get repeat demand. Marketplaces with better search and better spam resistance win users, while meetups that consistently bring serious traders together become nodes of economic value. Over time these practices harden into norms, and norms harden into institutions.

That process is slower than deploying software. It is also how markets have always grown.

24.7 Present Limits and Honest Boundaries

The institutional layer is still being built.

Reputation creates one obvious danger. The same continuity that lets others trust you also lets others track you: pseudonymity is workable because it carries history, and that same history can expose patterns and counterparties to anyone who watches long enough. A merchant may want one reputation for public commerce and a different one for sensitive trade, and a developer may want one identity for software and another for political writing. Identity separation remains hard in practice, because success creates the visibility it was meant to avoid.

Escrow creates another danger. The trusted third party may be narrower and less invasive than a state-regulated platform, but it can still become a chokepoint: an escrow agent with enough volume attracts coercion, and a well-known arbiter can be leaned on or made into a bottleneck for the commerce flowing through them. Decentralization lowers these risks without erasing them.

Physical goods remain difficult. Cryptography protects messages, but it does not hide the destination printed on a package or the schedule on which a courier arrives. Local pickup, lockers, drop points, and forwarding services reduce exposure in some cases, though none of them gives matter the freedom Tor gives packets. Digital commerce can become private faster than physical commerce ever will.

Adjudication is still thinner than settlement. Fund movement under pseudonymous arrangements already outperforms ordinary finance on trust minimization; complex dispute resolution has not yet reached comparable maturity. That gap is real.

Another gap waits beyond it. Money moves value through space, while credit moves value through time. A market that handles spot trade still needs ways to finance production and inventory. Bills of exchange, trade credit, bonded lending, and other forms of commercial trust belong to the next layer of buildout, and the tools for payment and communication have outrun the institutions for financing ongoing enterprise.[16]

These limits do not weaken the argument of the book. They give it shape.

The architecture exists for people to start moving economic life outside systems built for observation and control. The tools are here. The first institutions have appeared, and the remaining gaps are visible enough to work on.

That is enough to proceed with bounded trade and institutional experimentation.

The next chapter returns to the book's largest claim. What happens when these tools and institutions stop being isolated practices and begin to compound into a real economy?

Chapter Summary

Tools solve transport and settlement; they do not settle reliability or the judgment problems that arise when strangers trade under pseudonyms. A parallel economy needs institutional structure alongside its technical stack, and reputation under pseudonymity is the starting point. Persistent pseudonyms can accumulate trust while refusing civil identity, and the cost of abandoning accumulated reputation disciplines behavior. The tension is that the same continuity that enables trust also enables tracking. No deployed system maximizes both reputation and unlinkability at once, and reputation systems can reconcentrate into dossier infrastructure if every trade demands signed proof and each proof gets stored permanently. The discipline required is selective disclosure: reveal what the transaction needs and no more. Zero-knowledge credentials and attestation-based reputation protocols, still immature, open the prospect of separating reputational claims from the specific transactions and identities that produced them.

Escrow, multisignature arrangements, staged commitments, and release conditions make fraud costly and honest trade easier for most bounded disputes. These mechanisms reallocate trust instead of removing it: the parties still trust the keys and the software, and some arrangements also require trust in an arbiter, but narrow trust is often enough. Harder cases involving contested quality or ambiguous performance require arbitration structures that remain immature relative to state courts. The parallel economy has solved

settlement more cleanly than it has solved adjudication, and private arbitration has not reached the procedural depth long practice builds. Institutional emergence follows use: merchants with good dispute practices attract repeat business, and meetups that bring serious traders together become economic infrastructure.

Proxy merchants belong to the same pattern. Exchangers and forwarding operators emerge because they solve a real bottleneck at the edge of the market, and they remain healthy only while their role stays competitive and replaceable.

The institutional layer is not complete. Credit and trade finance remain largely unbuilt for pseudonymous commerce, and the tools for payment and communication have outrun the institutions for financing ongoing enterprise. Physical goods remain constrained by the anonymity gap that digital tools cannot bridge. These limits give the project shape without undermining its direction. The architecture exists for bounded trade and institutional experimentation to proceed, and Chapter 25 returns to the larger question of what happens when these practices compound.

Notes

[1]Timothy C. May, *The Cyphernomicon* (September 10, 1994), the comprehensive FAQ document May circulated on the Cypherpunks mailing list covering cryptography, anonymity, digital cash, and the political implications of strong encryption, `https://nakamotoinstitute.org/library/cyphernomicon/`. The epigraph is drawn from section 12.3. For the shorter programmatic statements, see May, "The Crypto Anarchist Manifesto" (1988), `https://nakamotoinstitute.org/library/crypto-anarchist-manifesto/`, and "Crypto Anarchy and Virtual Communities" (1994), `https://nakamotoinstitute.org/library/virtual-communities/`.

[2]For readers who want to work through the literature on private institutions and polycentric governance, Elinor Ostrom, *Governing the Commons* (Cambridge, 1990) is the canonical empirical study of self-organized institutions solving coordination problems without state enforcement. Bruce Benson, *The Enterprise of Law* (Pacific Research Institute, 1990) and *The Law* in *To Serve and Protect* (NYU Press, 1998) cover historical examples of private law and private policing. Edward Stringham, *Private Governance* (Oxford, 2015) applies the same approach to modern financial and online systems. For the theoretical side, David Friedman, *The Machinery of Freedom*, 3rd ed. (2014) is the most readable case for polycentric

law; Tom W. Bell, *Your Next Government?* (Cambridge, 2017) covers contemporary jurisdictional innovation. On reputation and commerce under weak enforcement, Paul Milgrom, Douglass North, and Barry Weingast, "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs," *Economics and Politics* 2, no. 1 (1990) is the key paper. For the longer civilizational framing, James Dale Davidson and William Rees-Mogg, *The Sovereign Individual* (1997), though dated in detail, remains the most ambitious attempt to describe the institutional implications of strong cryptography.

[3]Eric Hughes, "A Cypherpunk's Manifesto" (1993), `https://www.activism.net/cypherpunk/manifesto.html`: "Since we desire privacy, we must ensure that each party to a transaction have knowledge only of that which is directly necessary for that transaction."

[4]Nick Szabo, "Formalizing and Securing Relationships on Public Networks" (1997), `https://nakamotoinstitute.org/library/formalizing-securing-relationships/`. Szabo identifies search, negotiation, commitment, performance, and adjudication as the phases that smart contracts and digital institutions must address.

[5]Nick Szabo, "Negative Reputations" (1996), `https://nakamotoinstitute.org/library/negative-reputations/`. Szabo isolates the tension the main text describes: local nyms preserve privity, but once a pseudonym accumulates too much negative information it can be discarded for a fresh one unless counterparties demand stronger continuity or non-negative credentials.

[6]Jan Camenisch and Anna Lysyanskaya, "An Efficient System for Non-Transferable Anonymous Credentials with Optional Anonymity Revocation," EUROCRYPT 2001, is the foundational paper for anonymous credentials; BBS+ signatures and the W3C Verifiable Credentials standard carry the idea into more recent practice. Semaphore, maintained by Ethereum Privacy and Scaling Explorations (`https://semaphore.pse.dev`), provides anonymous group-membership proofs suitable for anti-sybil signaling and has been used in projects including Zupass, Worldcoin, and Bandada. UniRep (`https://developer.unirep.io`) is an attestation-based protocol specifically targeting reputation: attesters sign statements about a user, the user later proves aggregate properties under zero knowledge, and linkability between the proof and the underlying attestation history is broken. These systems remain research-grade or early production, not yet deployed marketplace infrastructure, but the primitives are established.

[7]The framework Chapter 6 develops rules out reputation as a property entitlement. What May, Hughes, and Szabo each name is the empirical role reputation plays in anonymous business: counterparties form decisions based on signed attestations and consistent behavior over time, not on any property right in being well-thought-of. The reputation systems this chapter describes carry empirical signals; they do not implement ownership claims over what others believe. See Chapter 6, §6.5; Chapter 21, §21.5; and Stephan Kinsella, "Defamation as a Type of Intellectual Property," in *A Life in Liberty* (Mises Institute, 2024).

[8]Nostr Improvement Proposal 05 (Mapping Nostr keys to DNS-based internet identifiers) defines the DNS- and HTTPS-backed identifier mapping used for human-readable Nostr names, https://github.com/nostr-protocol/nips/blob/master/05.md.

[9]Nostr already contains protocol pieces useful for challenge-response proof of key control. NIP-42 standardizes relay authentication challenges (https://github.com/nostr-protocol/nips/blob/master/42.md), while NIP-59 defines gift wrapping: a three-layer scheme of rumor, seal, and gift wrap that obscures message metadata including sender, recipient, and timestamp (https://github.com/nostr-protocol/nips/blob/master/59.md). NIP-59 provides privacy for message routing but is not identity proof by itself.

[10]Selective-disclosure tooling has reached production scale on both community and state sides: Semaphore and Zupass at conference-scale events, zkLogin for OAuth-derived wallets on Sui, W3C Verifiable Credentials as the generic format, ISO/IEC 18013-5 mobile driver's licenses in several countries; whether the tooling delivers the privacy it encodes depends on verifier-side posture, not protocol feature set. Semaphore, https://semaphore.pse.dev/. Zupass, https://zupass.org/. zkLogin on Sui, https://docs.sui.io/concepts/cryptography/zklogin. W3C Verifiable Credentials Data Model, https://www.w3.org/TR/vc-data-model/. ISO/IEC 18013-5:2021, "Personal identification, ISO-compliant driving licence, Part 5: Mobile driving licence (mDL) application," https://www.iso.org/standard/69084.html. Apple Wallet driver's license support and state rollout at https://support.apple.com/guide/wallet/use-your-drivers-license-or-state-id-apd1b8b6b87e/web. EU Digital Identity Wallet reference implementation at https://github.com/eu-digital-identity-wallet. On the verifier-side incentive problem, see Center for Democracy & Technology, "The Limits of Zero-Knowledge Proofs as a Privacy Solution," https://cdt.org/.

[11]Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" (2008), https://bitcoin.org/bitcoin.pdf. Section 4 describes the chain of digital signatures that constitutes ownership in the Bitcoin protocol; the multisignature extension—enabling M-of-N key arrangements—was introduced via the OP_CHECKMULTISIG opcode in the original codebase and later standardized in P2SH (BIP 16, January 2012) and P2WSH (BIP 141, November 2015). For the BIP 16 specification, see https://github.com/bitcoin/bips/blob/master/bip-0016.mediawiki. For BIP 141, see https://github.com/bitcoin/bips/blob/master/bip-0141.mediawiki. The two-of-three arrangement described in the main text is the structure Bisq uses for its escrow mechanism: https://bisq.network/faq.

[12]Nick Szabo, "Rights, Remedies, and Security Models" (2004), https://nakamotoinstitute.org/library/rights-remedies-and-security-models/. Szabo's point is that digital systems can enforce only a

narrow subset of what courts handle in the full range of disputes, because computer security relies on self-enforcing mechanisms and logs while law relies on remedies, procedure, and force behind judgment. That distinction is the one this section is drawing.

[13]Stephan Kinsella, Alessandro Fusillo, David Dürr, and Patrick C. Tinsley (Hans-Hermann Hoppe, adviser), *The Universal Principles of Liberty* (August 14, 2025), `https://stephankinsella.com/principles/`, is one such effort. It articulates a systematic libertarian private-law baseline including non-aggression, self-ownership, original appropriation, voluntary exchange, and rectification, alongside an Article IV on "Decentralised Legal Order" that explicitly endorses competitive arbitration and customary evolution of law without coercive monopoly. The Scope and Hierarchy section is explicit that adopters may build Secondary Codes and Laws on this baseline, with disputes over incompatibilities resolved by impartial arbitration. The document was itself inscribed on the Bitcoin blockchain, which illustrates how cryptographic rails can carry the institutional artifacts a parallel economy depends on.

[14]Kleros is the most widely deployed protocol-governed dispute resolution system: an Ethereum-based system in which disputes are submitted on-chain, jurors staked in PNK tokens are selected by Schelling-point coordination, and decisions are enforced through smart-contract fund release. Kleros, `https://kleros.io`; technical whitepaper: Clément Lesaege and Federico Ast, "Kleros: Short Paper v1.0.7," `https://kleros.io/whitepaper.pdf`. For the Aragon Court variant (now Aragon Govern), see `https://aragon.org`. An academic treatment of the Schelling-point coordination mechanism underlying token-curated juror selection is Thomas Schelling, *The Strategy of Conflict* (Harvard University Press, 1960), chapter 3. These systems handle narrow on-chain disputes reliably; their procedural depth for complex off-chain fact disputes remains limited by the jury's inability to inspect evidence outside what is submitted on-chain.

[15]On proxy merchants and adjacent boundary roles, see Smuggler and XYZ, *Second Realm: Book on Strategy* (Liberty Under Attack Publications, 2015), especially the sections on specialization, proxy merchants, shared services, and OTC exchangers. For the earlier transition logic behind the same move, see Samuel Edward Konkin III, *New Libertarian Manifesto* (1980). These sources are useful here because they name a real institutional function that the cryptographic stack alone does not name.

[16]On credit, trade finance, and the capital structure of production in a free market, see Murray N. Rothbard, *Man, Economy, and State*, 2nd ed. (Ludwig von Mises Institute, 2004), chapters 5–6, which analyze capital goods, stages of production, and the time-preference theory of interest without which credit cannot be understood. Jesús Huerta de Soto, *Money, Bank Credit, and Economic Cycles*, 3rd English ed. (Ludwig von Mises Institute, 2012), traces the historical development of credit instruments including bills of exchange and the institutional conditions under which commercial trust makes production financing possible. The gap identified at the end of §24.7—that payment and

communication tools have outrun the institutions for financing ongoing enterprise—maps precisely onto the distinction Rothbard draws between the money relation and the credit relation: money moves purchasing power through space, credit through time, and the institutional prerequisites for each are different.

25

BUILDING THE PARALLEL ECONOMY

> "We will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before."
>
> John Perry Barlow, "A Declaration of the Independence of Cyberspace" (1996)[1]

This book began with a challenge: "If you have nothing to hide, you have nothing to fear." It promised a complete answer that joined economics with cryptography, and philosophy with engineering.

The preceding chapters built the case in stages: foundations in Part II, exchange and sound money in Part III, the adversary in Part IV, and the tool stack in Part V, from cryptography and anonymous networks to Bitcoin, zero-knowledge proofs, and decentralized social infrastructure.

25.1 The Convergence of Foundations

The book's three axioms operate at different levels but point in the same direction.

The Action Axiom establishes that privacy is structural. Human action is purposeful behavior that requires internal deliberation, and the actor necessarily possesses information others lack: preferences and plans that exist in the mind before disclosure. That asymmetry is built into action as human beings deliberate and choose, which

means privacy exists as a descriptive fact before any normative claim is made about it.

The Argumentation Axiom establishes that privacy cannot be coherently denied. To argue at all is to exercise control over one's body and mind, presupposing the self-ownership from which privacy derives. The surveillance advocate who argues for surveillance shows through the act of arguing the autonomy he seeks to deny others. This point is stronger than persuasion alone. It identifies a performative contradiction inside anti-privacy discourse.

The Axiom of Resistance establishes that privacy can be technically defended. Computational hardness assumptions, tested by decades of failed attacks, provide foundations for systems that can raise the cost of control sharply under hostile conditions. Unlike the Action Axiom, this claim is empirical and not self-evident, but the record supports it: Tor operates, Bitcoin processes blocks, encrypted messages reach their destinations, and resistant systems have survived sustained pressure in practice.

Together, these axioms create a complete foundation. Privacy exists as fact. It also supports a defensible norm and remains technically possible under hostile conditions. The parallel economy emerges from this convergence: systems that implement what theory establishes and ethics requires.

25.2 The Synthesis

The book's distinctive contribution is what emerges when the components combine.

Chapter 2 established that praxeology and cypherpunk cryptography, developing independently, converged on the same conclusions about privacy, spontaneous order, and sound money. That convergence suggested both traditions discovered something true. This chapter asks: what do their combined insights enable?

The technical stack exhibits reinforcing properties. Each tool addresses one vulnerability, and together they support forms of trade and coordination that were previously fragile or costly. A merchant can receive payment in Bitcoin through Tor, communicate with

customers through encrypted channels, prove credentials without revealing identity, and maintain reputation through Nostr with less dependence on the surveilled financial system. These pieces already function, even if the full stack remains uneven and incomplete.

This synthesis arrives at a particular historical moment in which the tools are finally mature. Bitcoin has run without interruption for more than fifteen years; Tor has operated for more than twenty; end-to-end encryption ships in the messaging applications ordinary people use every day. The threat is intensifying in parallel: CBDCs advancing toward deployment, surveillance infrastructure expanding, regulatory pressure on exchanges and infrastructure providers increasing. Earlier cypherpunks had vision without infrastructure; earlier Austrians had theory without implementation. The present moment offers both at once.

25.3 Breaking the Observation Loop

Taxation and confiscation both require the ability to identify wealth and compel its surrender. Inflation operates the same way at the monetary layer: debasing a currency requires that the currency be the only medium in circulation. When identification becomes less reliable or more expensive, transfer becomes harder and costlier. One transaction escaping observation is a small effect; many transactions escaping routine observation makes revenue collection less efficient and control less precise. The state does not need to vanish for this shift to carry weight; surveillance becoming a weaker and more expensive basis for rule is enough.

CBDCs are the state's counter-move: restructuring money itself so that observation is automatic and unavoidable, because if all transactions occur through state-controlled infrastructure the Observe stage becomes trivial. The parallel economy is the counter-counter-move. Every transaction outside surveillance infrastructure shows that observation is not inevitable. The race is between surveillance infrastructure that keeps control cheap and privacy infrastructure that makes control more costly and less reliable.

25.4 What Has Been Achieved

The cryptoanarchist vision articulated in Chapter 2 has partially materialized. The record begins with what works.

Bitcoin has run without interruption since January 2009. No government has eliminated it despite sustained attempts: when China banned Bitcoin mining in 2021,[2] the network's hashrate recovered within months as mining redistributed across jurisdictions. Consensus has continued under sustained adversarial pressure, which is the operational evidence the theoretical chapters predicted.

Tor routes substantial user traffic through a network no single entity controls. Hidden services host markets and communication channels that persist despite law enforcement operations. When services are seized, replacements emerge. The architecture absorbs attacks and continues operating.

Encrypted messaging has achieved deployment that would have seemed fantastical to 1990s cypherpunks. Signal is now a mainstream global application, and end-to-end encryption ships at enormous scale across consumer messaging. The "going dark" problem that law enforcement laments is evidence of success: communications that cannot be intercepted even with lawful authority are exactly what the cypherpunks intended.[3]

The theoretical framework predicted this outcome. Praxeology holds that spontaneous order emerges from voluntary exchange without central planning, and the parallel economy shows the principle at work; the Axiom of Resistance holds that properly designed systems can resist control, and the record of the past decade and a half supports that claim. Theory and evidence meet at an operational threshold neither side had reached on its own.

Parallel Institutions in the Same Territory

The parallel economy grows inside the same cities and commercial districts where the official economy already operates, without waiting for a separate geography. The private workshop sits down the street from the regulated office park. The Bitcoin meetup happens a few blocks from the bank branch. A merchant who accepts

bitcoin and coordinates through encrypted channels may still rent space in a city fully mapped by the state.

This changes what success looks like. Success is the growth of parallel institutions that can meet, trade, teach, and settle without asking permission from the old rails every time they move, not territorial purity. The digital stack made this possible; physical bridgeheads make it durable.

Hakim Bey supplied the early language of the temporary autonomous zone. The practical form that concerns this book is narrower and more durable. It means recurring pockets of local autonomy: meetups, workshops, merchant rooms, and trade nodes embedded in ordinary cities while more of their coordination moves onto private rails.[4]

The Cryptographic Stack Has Matured

The primitives examined in Parts IV and V have moved from research into production within the time this book was being written. Zero-knowledge proofs that required trusted setups a decade ago now run on recursive constructions that eliminate the setup entirely, and systems including Semaphore, Zupass, and Sui's zkLogin carry them into consumer-scale event ticketing, credentialing, and OAuth-based wallet derivation. Fully homomorphic encryption is now deployed in narrow applications after four orders of magnitude of efficiency improvement since Gentry's 2009 construction.[5] Multi-party computation secures roughly two thousand institutional custody customers through threshold signing, and the Schnorr-based threshold schemes MuSig2 and FROST make n-of-n and t-of-n Bitcoin multisig indistinguishable from single-sig transactions on-chain. Trusted execution environments sit in the hardware of hundreds of millions of consumer devices and now back cloud-scale attested services through Apple Private Cloud Compute, Azure Confidential Computing, AWS Nitro Enclaves, and Google Cloud Confidential Computing. Differential privacy protects the 2020 U.S. Census release and the telemetry of every recent iOS device. Private information retrieval has reached practical deployment in Signal's contact discovery and in open implementations around SealPIR and

SimplePIR. Post-quantum migration has begun in Signal's PQXDH protocol, Apple iMessage PQ3, Chrome's hybrid key exchange, and Cloudflare's edge, with NIST's lattice and hash-based standards finalized and a Bitcoin migration proposal (BIP-360's Pay-to-Merkle-Root) in active specification. Monero's FCMP++ upgrade, designed to replace ring signatures with chain-wide membership proofs, is in preparation.

The transport layer has expanded beyond the ordinary internet. Meshtastic deployments have grown at large events where the conventional network is saturated or untrusted. Reticulum decouples cryptographic addressing from any specific physical medium. The Free Internetworking Peering System (FIPS) provides a transport-agnostic overlay whose routing addresses are self-generated cryptographic identities, independent of any ISP, registrar, or certificate authority. BitChat brought Bluetooth-mesh messaging to a mass-market audience. Satellite constellations including Iridium, OneWeb, Starlink, and Globalstar carry traffic that does not depend on any single state's terrestrial infrastructure.

The adversary has matured alongside the defense. Chapter 12 described the Analytics Stack the state now buys from commercial suppliers: sensor networks, location brokers, AI-driven data-fusion platforms, commercial spyware, and attested decision systems that compress the entire OODA loop into an industrial process. The defender's response moves with the adversary. Content encryption stays essential and no longer suffices, because a metadata-capable model reaches conclusions from timing, social graph, location, and frequency when content is unavailable. Transport anonymity, mesh routing, private information retrieval, and computing-on-secrets are the operational answers to the metadata-analyst side of the stack, and their maturity is the reason this book treats them as first-tier infrastructure.

The stack described here did not exist in the form Chapter 2's original cypherpunk vision imagined. It exists now. The theoretical architecture the book traced across the preceding chapters is what the world has already built.[6]

The Empirical Record and What It Tests

The parallel economy is the book's central operational claim, and any operational claim must be testable against what the world produces. The record accumulated across the last decade is partial, and the partiality itself is informative. Each case confirms part of the theory and illustrates specific structural pressures the earlier chapters named in the abstract.

Circular Bitcoin economies have emerged where two conditions held together: a permissive local regulatory environment and a commercially entrepreneurial community committed to running the experiment long enough to reach a density threshold. Where those conditions held, merchants accepted bitcoin, wages and savings denominated in it, all without a state monetary authority operating in the medium. Where any one of those conditions did not hold, the experiments did not reach density. The theory predicts this. Spontaneous order emerges from voluntary exchange at whatever rate the participants can sustain, and the participants' willingness to sustain the exchange is the binding constraint, not any technical property of the money.

El Salvador's reversal of Bitcoin's legal-tender status in early 2025, under IMF financing pressure, is worth naming because it tests a specific secondary claim and leaves the primary one alone. The primary claim of this book is that parallel institutions can be built on cryptographic rails and can operate without the state's permission. The secondary claim, sometimes made by enthusiasts, is that state adoption of Bitcoin is itself a meaningful step toward the parallel economy. The reversal refutes the secondary claim without touching the primary one. A state that adopts Bitcoin under domestic enthusiasm and then abandons that status under external financing pressure has shown that the adoption was not a structural transition. The structural transition happens in the voluntary economy of merchants and holders who act without the state's blessing, and in that economy the circular activity in Bitcoin Beach, El Zonte, and the surrounding communities continued after the legal-tender designation was withdrawn. What the state declared legal tender,

it could withdraw. What participants chose to trade, it could not unchoose.[7]

A second empirical pattern is the gentrification effect observable in several early circular-economy locations. When a small coastal town becomes a destination for Bitcoin-wealthy visitors, land prices rise sharply and the local population that the experiment was meant to serve can find itself priced out of its own neighborhood. This is the general pattern of any locality that becomes attractive to wealthier outside participants, visible in university towns and resort villages across every country for a century. The Bitcoin case adds a new variable to an old phenomenon. Parallel economies, like any economies, have distributional consequences, and the first-order benefits of the new monetary architecture do not automatically accrue to the participants the early rhetoric named.

A third pattern is the fragility of centralized peer-to-peer marketplaces. Successive waves of services that brokered no-KYC bitcoin-for-cash trades have closed under regulatory pressure or for reasons internal to their operators, while the decentralized alternatives that survived, including Bisq and RoboSats, had architectures in which no single operator could be forced to shut the service down. The pattern confirms the theoretical point the book has made repeatedly. Decentralization protects systems more reliably than it protects the people who use them, and the people who run centralized services against the state's preference remain the system's fragile point.[8]

These empirical observations do not weaken the book's thesis. They sharpen it. The parallel economy is operational, and its operation is uneven. The unevenness is predictable: protocol-level decentralization lets systems survive pressure that destroys operator-dependent services; cultural commitment to sustaining the practice lets circular economies take root where mere technical access does not; and the distributional pressures that first-order benefits attract are problems the next iteration of the project must anticipate and address.

The right reading of the record is neither triumphalist nor deflationary. It is the reading the Axiom of Resistance already suggests: a system's value is proportional to the cost required to compromise it,

and the cost is measured empirically by what specific pressures have accomplished against specific deployments. The deployments have absorbed considerable pressure. They have also revealed specific points of fragility that the next phase of engineering and institution-building will have to address. Both halves of that sentence are equally important, and both are what this section establishes.

25.5 Limits and Open Questions

The parallel economy is real but constrained. The constraints that follow shape what adoption can realistically achieve today, and each outweighs any general claim of success.

Scale and performance constraints

Bitcoin, Lightning, Tor, and Signal all operate at meaningful scale in absolute terms and at small scale relative to global finance and communication. The point is not dominance but viability: each system has shown that the core architecture can function under real adversarial conditions, which is what the theoretical chapters predicted and the question the empirical record needed to answer.

Scale still binds in several concrete ways. Bitcoin block space remains scarce, and fee pressure during periods of high demand can make small transactions economically irrational on the base layer. Lightning increases payment frequency substantially but still depends on on-chain entry and exit, and channel management requires sustained attention that most users do not provide. Tor runs slower and less reliably than centralized alternatives because onion routing through volunteer relays adds latency that no optimization can eliminate without sacrificing the anonymity property the routing provides. Privacy tools consistently demand more effort than surveilled alternatives because the overhead that produces privacy is structural: anonymity sets require participants, zero-knowledge proofs require computation, and distributed consensus requires coordination, while centralized surveillance services pay none of these costs and capture the efficiency gain as a product advantage.

The Reputation-Anonymity Tension

Chapter 24 laid out this tension in detail. The summary here is brief because the argument is already made: the tradeoff is real and structural, not a temporary deficiency waiting for a better implementation. A pseudonym that accumulates trust does so by being consistent and persistent, which means its behavior is linkable across contexts. A fresh key that preserves unlinkability carries no history, which means counterparties have no basis for extending trust beyond the minimum the immediate interaction requires. Zero-knowledge attestation schemes offer a partial escape by separating reputational claims from the identities that produced them, but deployment remains early and the schemes still depend on attesters who observe the underlying facts. The tension shapes what the parallel economy can accomplish in any given context and determines how much friction reputation-dependent commerce carries until the tooling matures.

Physical Goods and the Anonymity Gap

Digital privacy tools protect digital activity. Physical goods create an anonymity gap that technology cannot bridge.

Shipping requires physical addresses. No amount of encryption protects the destination printed on a package; customs inspection exposes contents and couriers track delivery. The "last mile" problem is physical, not technical: goods must arrive somewhere, and that somewhere can be observed.

Partial mitigations exist. Receiving at neutral locations such as PO boxes or package lockers shifts exposure from home address to pickup location, and remailers and forwarding services add intermediaries that blunt the observation. Each adds delay and new failure points without closing the gap, and none achieves for physical goods what Tor achieves for digital traffic.

Academic work on anonymous physical delivery exists (APOD, Lelantos) but has not achieved practical deployment.[9] The problem is not purely technical; physical objects cannot be copied and

rerouted like data packets. The physics of matter constrains what cryptography can achieve.

This constraint was observed and measured by agorist theorists before cryptographic markets existed. Writing in 1973, Rayo (Tom Marshall) formalized the relationship under the banner of vonu (a contraction of "voluntary not vulnerable") and plotted what he called Mean Time to Harassment against an activity axis running from summer survival through all-weather survival, comfortable home, small workshop, small manufacturing, light industry, and heavy industry. Each order-of-magnitude increase in activity required roughly a tenfold increase in operational competency to hold time-to-detection constant, and his floor for a meaningful alternative economy was the small-workshop level.[10] The axes today are cryptographic as much as geographic, but the curve has not changed. Small-scale digital activity can be hidden for years with modest effort, while physical-scale production and distribution remains constrained by the same anonymity gap Rayo described half a century ago.

Economic activity includes physical goods. A parallel economy for digital services is possible today. A parallel economy built around physical production and distribution remains constrained by the anonymity gap.

Lightning Network Privacy Limitations

Chapter 20 presented Lightning Network as providing payment privacy superior to base-layer Bitcoin. This is true but requires qualification.

Lightning privacy is not absolute.[11] Channel opening and closing transactions are visible on the blockchain; these can reveal approximate capacity and associate channels with on-chain activity. If the bitcoins funding a channel are linked to an identity (typically through KYC exchanges), that linkage persists. Mobile wallets typically connect to Lightning Service Providers (LSPs) that see all the user's payment activity, shifting trust from the network to the LSP. Routing analysis can, in some cases, infer payment paths, especially for large payments with few viable routes.

Best practices mitigate these limitations: open channels with coinjoined funds and prefer self-hosted nodes over custodial services. But these practices require technical sophistication most users lack. Default Lightning usage provides much better privacy than default on-chain Bitcoin but falls short of ideal anonymity.

The Post-Quantum Transition

NIST finalized the core lattice and hash-based signature standards (ML-KEM, ML-DSA, SLH-DSA) in August 2024 and selected the code-based HQC backup KEM in March 2025.[12] Mainstream deployment has begun: Signal's PQXDH protocol since September 2023, Apple iMessage PQ3 since February 2024, Chrome's X25519MLKEM768 hybrid key exchange through 2024, and Cloudflare's edge carrying post-quantum TLS past a material fraction of handshakes by mid-2025. Bitcoin's migration lags, because any softfork proposal must address not only new spends but the treatment of pre-existing at-risk outputs.[13] The current working proposal is BIP-360's Pay-to-Merkle-Root (P2MR) output type, which is Taproot with the key-path spend removed; the post-quantum signature opcodes are a separate future layer on top of it. Neither has been activated.

Quantum-relevant computers remain behind the threshold Shor's algorithm requires at production scales, but harvest-now-decrypt-later makes the migration horizon tighter than the hardware timeline suggests. Bitcoin's migration must complete before a cryptanalytically relevant quantum computer appears; the activation path is the first substantive protocol-level decision since Taproot.

Dispute Resolution Without Courts

Chapter 24 carried the full treatment of private dispute resolution. The short version belongs here: bounded disputes already have workable tools. Escrow, multisignature arrangements, and staged commitments handle the large majority of market disputes, which are disputes over delivery, payment, and clear-rule violations, not

contested interpretations of complex fact. Hard disputes involving quality judgments, ambiguous performance, or contested testimony remain thinner than state courts because private arbitration lacks the procedural depth, precedent, and external enforcement authority that state courts have accumulated over centuries of practice. That gap is real, does not stop the parallel economy from growing through the much larger volume of clean disputes, and marks one clear frontier in the institutional buildout that sustained commercial activity will eventually require.

Mainstream Adoption Barriers

Privacy tools still require more technical knowledge than surveilled alternatives, and that gap persists even as individual tools improve. A new user setting up a self-custodial Lightning wallet, running a Tor circuit, and managing a Nostr keypair faces a configuration burden that a new user opening a bank account or downloading a surveilled messaging app does not. Network effects compound the problem: the value of any communication tool depends on whether the people you need to reach also use it, so privacy tools compete against platforms whose user bases already include everyone the new user knows. Surveilled services often provide a better experience within their own design constraints, because they can optimize aggressively for speed and convenience without paying the overhead that privacy properties cost.

Legal uncertainty creates a second barrier that technical improvement cannot resolve. Jurisdictions differ substantially on what is permitted: the same exchange, tool, or practice that is unremarkable in one country may be a regulatory violation in another. Regulatory pressure on exchanges and infrastructure providers can remove the on-ramps and off-ramps that make private financial tools usable for ordinary commerce. End users face uncertainty about whether their own use of privacy tools is lawful and about whether that legal status will change. The Tornado Cash prosecution, examined in Chapter 13, showed that legal risk reaches not only users but builders: writing and publishing code that the state later decides it dislikes can result

in criminal charges, and that chilling effect extends well beyond the specific developers prosecuted.

What These Limitations Mean

The limitations above do not constitute an objection to the project. They constitute its honest description. Most people still depend on state-supervised systems for most activity, and the parallel economy operates today as a supplement to the old system, not a replacement for it. That is where it is in 2026. Replacement, if it comes, is a later stage that follows from the present one: tools improve, communities grow, institutional infrastructure matures, and each marginal participant who moves some activity onto private rails makes the network slightly more useful for the next one.

The right frame is not whether the parallel economy has won but whether it provides value now and whether the direction of change is toward greater privacy and autonomy or away from it. On both counts the answer is yes. The tools described in this book process real transactions, protect real communications, and support real commerce under adversarial conditions that would have destroyed earlier attempts. The limitations are real, finite, and mapped. That is enough to proceed.

25.6 Answering ``Nothing to Hide''

Chapter 1 posed the challenge: "If you have nothing to hide, you have nothing to fear." This claim justifies surveillance by asserting that only wrongdoers need privacy. After the preceding analysis, the complete answer emerges.

At the structural level, the claim misunderstands what privacy is. Chapter 3 established that privacy is inherent to human action, because human action necessarily includes internal deliberation and subjective valuation under conditions of information asymmetry. Privacy is the space in which thought occurs and action is chosen, and every human being requires it whatever their acts.

At the informational level, the relevant question is not whether anyone is hiding wrongdoing but how information flow is controlled

at all. Chapter 7 established that exchange functions best when parties control disclosure. Knowledge of preferences and plans enables exploitation, and transparent negotiation collapses into advantage for the more desperate party. Privacy protects deliberation and enables voluntary exchange.

Surveillance distorts market processes at the economic level. Price signals degrade when participants fear observation, and capital flows follow regulatory visibility instead of economic merit. Economic calculation depends on information that surveillance compromises.

The political dimension is more direct. Surveillance enables control independent of prosecution. The opposition that can be monitored can be neutralized: donors identified and pressured, organizers tracked and harassed, plans discovered and preempted. Totalitarian regimes require surveillance not because all citizens are criminals but because surveillance enables control. "Nothing to hide" ignores that the watching itself is the threat.

The definitional problem runs deeper than any of the above. "Nothing to hide" presupposes a stable, knowable boundary between innocent and suspicious. But who defines "something to hide"? The state determines what is prohibited, and prohibitions change. Activity legal today may be criminalized tomorrow. The question is not whether you have something to hide now, but whether you might have something to hide under any future interpretation by any future authority.

Acknowledging that uncertainty does not mean privacy has no costs or that tradeoffs do not exist. Privacy tools can shelter wrongdoing. Anonymity can protect criminals alongside dissidents. The parallel economy operates outside the state's legal apparatus, which means outside its protections as well as outside its restrictions, and the private-law institutions that might substitute for state courts are, as Chapter 24 argued, still maturing. These costs are real and should not be dismissed.

Privacy carries costs. Surveillance distorts markets and enables control, threatening the conditions under which voluntary coordi-

nation functions. Those costs are larger and less reversible than the costs privacy imposes.

The answer to "nothing to hide" is plain. Privacy is not about hiding. It protects the conditions necessary for human action and for coordination through markets.

"Nothing to hide" inverts the burden of proof. In any system that respects human agency, the question is not "why do you need privacy?" but "by what right do you demand access?" Self-ownership means that what one thinks and what one does are private by default. Intrusion requires justification, and withholding does not.

The cypherpunk answer adds a harder practical point: even if surveillance were claimed as legitimate, it need not remain cheap or routine at scale. Mathematics provides tools that can make many forms of surveillance unreliable and expensive. That does not erase the question of justification. It changes the terrain on which the question is fought.

25.7 Conclusion

This book has traced an argument from axiom to implementation, from theory to operational reality.

Privacy is structural to human action. It cannot be coherently denied in rational discourse. It can be defended through cryptographic tools that raise the cost of control and preserve room for voluntary coordination.

States surveil because observation enables targeting and collection, and because both support control. When observation becomes unreliable or expensive, those mechanisms weaken. The apparatus of financial surveillance and identity requirements, together with broader regulatory control, depends on keeping economic life legible enough to monitor and interrupt.

The mathematics have not abolished conflict, but they have changed its economics. Defense can be cheap while attack becomes expensive. A transaction hidden from routine observation is harder to tax; a wallet that cannot be confidently linked is harder to seize.

When these costs rise across enough activity, control loses efficiency and reach.

The parallel economy already processes transactions the state did not authorize and routes messages agencies cannot routinely read. Storing value outside ordinary seizure channels and encrypting default communication channels remove activity from routine observation. As adoption spreads, these gains compound.

The state claims a monopoly over money and communication, and therefore over the coordination built on both. That claim is already weaker than it appears. Bitcoin operates, Tor keeps routing, Signal keeps delivering, the computing-on-secrets primitives work at the scales they were built for, and the post-quantum migration has begun in production systems. The monopoly persists where people still depend on rails built for surveillance and control, and it retreats where they do not.

The work is practical: run a node, generate keys, encrypt by default, transact privately, join or start a meetup, contribute to an open implementation, and build the habits and institutions that let the tools reinforce one another. The tools exist, and their reach depends on whether people use them.

Build.

Notes

[1]John Perry Barlow, "A Declaration of the Independence of Cyberspace," delivered in Davos, Switzerland, February 8, 1996, `https://www.eff.org/cyberspace-independence`. Barlow co-founded the Electronic Frontier Foundation and wrote the Declaration in response to the 1996 Telecommunications Act. The closing lines quoted above frame the parallel-economy argument this chapter makes: that a civilization built on different rails is not only possible but is already being built, whatever governments do. For the companion cypherpunk-movement statement, see Eric Hughes, "A Cypherpunk's Manifesto" (1993), `https://www.activism.net/cypherpunk/manifesto.html`.

[2]China's State Council announced the ban on Bitcoin mining in May 2021, with enforcement through June–July 2021 resulting in the expulsion of approximately 50–65% of global hashrate within weeks. The network's hashrate fell from roughly 180 EH/s in May 2021 to a low near 84 EH/s in July 2021 before recovering to prior levels by January 2022 as mining relocated to the United States, Kazakhstan,

Russia, and Canada. Coverage: *Cambridge Centre for Alternative Finance*, Cambridge Bitcoin Electricity Consumption Index hashrate data, `https://ccaf.io/cbnsi/cbeci`; Wolfie Zhao, "Bitcoin Hashrate Recovers to Pre-China Ban Levels," *The Block* (January 2022); Nic Carter, "How Much of Bitcoin Mining Uses Renewable Energy?" *CoinDesk* (2021).

[3]The "going dark" framing for the law enforcement challenge posed by end-to-end encryption was articulated formally by FBI Director James Comey before the Senate Judiciary Committee (July 8, 2015) and has been repeated by successive FBI and DOJ leadership. Signal's global deployment is tracked at `https://signal.org`; as of 2024 Signal reported over 40 million monthly active users. For the encryption policy debate, see Riana Pfefferkorn, "The 'Going Dark' Problem," Center for Internet and Society, Stanford Law School (2021), `https://cyberlaw.stanford.edu/`; and Susan Landau, *Listening In: Cybersecurity in an Insecure Age* (Yale University Press, 2017), which remains the most technically grounded treatment of the wiretap access debate.

[4]Hakim Bey, *T.A.Z.: The Temporary Autonomous Zone, Ontological Anarchy, Poetic Terrorism* (Autonomedia, 1991), supplied the phrase. For a later strategy text that translates the idea into recurring physical-and-digital parallel institutions, see Smuggler and XYZ, *Second Realm: Book on Strategy* (Liberty Under Attack Publications, 2015). The usage in this chapter is narrower than either text's rhetoric. It refers to durable local nodes of repeated trade and coordination inside the same territory as the official order.

[5]Fully homomorphic encryption (FHE) was first proved possible by Craig Gentry, "A Fully Homomorphic Encryption Scheme," Stanford PhD dissertation (2009), available at `https://crypto.stanford.edu/craig/craig-thesis.pdf`. The four-orders-of-magnitude efficiency improvement referenced in the main text reflects the progression from Gentry's original lattice-based scheme through the BGV, BFV, CKKS, and TFHE lineage; a useful survey is Ilaria Chillotti, Nicolas Gama, Mariya Georgieva, and Malika Izabachène, "TFHE: Fast Bootstrapping over the Torus," *Journal of Cryptology* 33 (2020), `https://eprint.iacr.org/2018/421`. Current production deployments include Zama's Concrete and TFHE-rs libraries (`https://github.com/zama-ai/tfhe-rs`) used in on-chain FHE smart contracts, and Intel HEXL for hardware-accelerated HE primitives.

[6]Chapter-by-chapter cross-reference to the new primitives summarized in this section: zero-knowledge proofs in Chapter 15 (notes 10, 11); threshold signing and MPC in Chapter 16 (note 4); trusted execution environments in Chapter 16 (note 5); attested confidential LLM inference in Chapter 16 (note 6); private information retrieval in Chapter 16 (note 7); differential privacy in Chapter 16 (note 8); post-quantum signatures in Chapter 14 (note 10) and Chapter 20 (note 17); Monero FCMP++ in Chapter 20 (note 15); FIPS and the mesh-transport stack in Chapter 17 (notes 17–22); the Analytics Stack adversary picture in Chapter 12 (notes 1–9). These notes carry the detailed references; this section summarizes what the stack now amounts to taken as a whole.

[7]El Salvador adopted Bitcoin as legal tender in September 2021 under the Bukele government and reversed the legal-tender designation in January 2025 as part of an IMF financing agreement; circular-economy activity (merchants, tourism, local Bitcoin payments) continued after the reversal, showing that state adoption and voluntary adoption are separable. El Salvador Ley Bitcoin, Decree 57 (June 8, 2021); reversal in the Legislative Assembly vote on January 29, 2025, covered in *Reuters* and *Bloomberg*. IMF Extended Fund Facility for El Salvador (approved December 2024), `https://www.imf.org/en/Countries/SLV`. On circular-economy continuity, see Bitcoin Beach documentation at `https://www.bitcoinbeach.com/` and reporting by *Forbes* contributor Francis Pouliot and on-the-ground coverage from *El Salvador in English* (`https://elsalvadorinenglish.com/`). On the wider El Salvador Bitcoin experiment, see Nicolás Cachanosky, Bryan Cutsinger, and Alexander Salter, "Bukelenomics: Radical Policymaking and Economic Development in El Salvador," *Journal of Private Enterprise* (2023), and the Human Rights Foundation's ongoing coverage at `https://hrf.org/`.

[8]Bisq and RoboSats are the decentralized no-KYC Bitcoin marketplaces that have outlasted successive waves of centralized-p2p service closures; architectural decentralization protected the systems where it could not protect operators of centralized alternatives. Bisq, `https://bisq.network/`, with source at `https://github.com/bisq-network/bisq`. RoboSats, `https://learn.robosats.com/`, with source at `https://github.com/RoboSats/robosats`. For the historical record of closed centralized p2p services including LocalBitcoins (wound down in February 2023) and Paxful (regulatory exit), see public announcements at the respective company pages and coverage in *CoinDesk* and *Bitcoin Magazine*. On the Tornado Cash prosecution as the adjacent pattern for protocol vs. operator distinction, see U.S. Department of Justice indictment at `https://www.justice.gov/usao-sdny/pr/tornado-cash-founders-charged-money-laundering-and-sanctions-violations`. Chapter 13 develops the protocol-versus-builder analysis of this pattern.

[9]Academic work on anonymous physical delivery (APOD, Lelantos) is documented in Ulrich Flegel, *Privacy-Respecting Intrusion Detection* (Springer, 2007); Giulia Fanti and Pramod Viswanath, "Anonymity Properties of the Bitcoin P2P Network" (2017); and George Danezis and Claudia Diaz, "A Survey of Anonymous Communication Channels," Microsoft Research Technical Report MSR-TR-2008-35 (2008), `https://www.microsoft.com/en-us/research/publication/a-survey-of-anonymous-communication-channels/`.

[10]Rayo (Tom Marshall), "Vonu: Mean-Time to Harassment," *VonuLife* (1973); reprinted in *Vonu: The Search for Personal Freedom*, ed. Shane Radliff (Liberty Under Attack Publications, 2020). Rayo plotted MTH on a vertical axis whose units each represented a tenfold increase in years of invisibility, against an activity axis scaled from summer survival through all-weather survival, comfortable home, small workshop or laboratory, small manufacturing, light industry, and heavy industry, with diagonal lines of operational competency spaced one order

of magnitude apart. His explicit conclusion was that "E-level [small workshop] is probably minimum for development of much of an alternative economy worthy of the name." Vonu occupied the same 1970s libertarian milieu as the early agorist and counter-economic thought later systematized by Samuel Edward Konkin III, and the metric translates cleanly into cryptographic markets: both measure freedom as a time-to-detection quantity against a coercive observer, differing only in whether concealment is achieved geographically or cryptographically.

[11]For a full analysis of Lightning Network privacy limitations, see lightningprivacy.com, particularly "Routing Analysis" and "Introduction to Lightning Privacy." Academic treatments include Nisslmueller et al., "A Cryptoeconomic Traffic Analysis of Bitcoin's Lightning Network" (2019), and Romiti et al., "An Empirical Analysis of Privacy in the Lightning Network," *Financial Cryptography* (2021).

[12]For CRQC (Cryptographically Relevant Quantum Computer) timeline estimates, see PostQuantum.com, "Q-Day Predictions: Anticipating the Arrival of CRQC" and "Q-Day Revisited: RSA-2048 Broken by 2030," both at `https://postquantum.com`. See also a16z crypto, "Quantum Computing and Blockchains: Matching Urgency to Actual Threats" (2024), `https://a16zcrypto.com`. NIST's post-quantum cryptography standards, particularly FIPS 203 (ML-KEM / Kyber), FIPS 204 (ML-DSA / Dilithium), and FIPS 205 (SLH-DSA / SPHINCS+), finalized August 2024, are at `https://csrc.nist.gov/projects/post-quantum-cryptography`. HQC (Hamming Quasi-Cyclic), a code-based KEM selected by NIST in March 2025 as a backup to the lattice-based ML-KEM, is documented in NIST IR 8545 and at `https://csrc.nist.gov/projects/post-quantum-cryptography`. For a broader academic survey, Bernstein, Buchmann, and Dahmen, eds., *Post-Quantum Cryptography* (Springer, 2009), remains the standard reference on lattice, code-based, hash-based, and multivariate constructions.

[13]Project Eleven, "Quantum Vulnerability of Bitcoin Addresses" (2025), `https://web.archive.org/web/20250501000000*/https://project11.xyz`, estimates approximately 6.5 million BTC held in addresses with exposed public keys. See also River, "Will Quantum Computing Break Bitcoin?" `https://river.com/learn/will-quantum-computing-break-bitcoin/`. Roughly 1.9 million BTC are in P2PK addresses (early 2009–2011 coinbase outputs) and another 4+ million BTC are in addresses whose public key was exposed by reuse after spending. For Bitcoin-specific post-quantum discussion, see the long-running bitcoin-dev mailing list threads on taproot-compatible quantum-resistant signatures, and the QuBit soft-fork proposal drafted by Hunter Beast and others at `https://github.com/cryptoquick/bips`.

ABOUT THIS EDITION

This is the *extended* edition of *The Praxeology of Privacy*, prepared for digital and e-reader formats. The endnotes throughout this volume carry the author's full annotated commentary: bibliographic essays on each cited source, reading lists for the topics each chapter develops, and pointers to the working tools and current standards referenced in the technical chapters.

A companion print edition is also available. To keep the printed book within a portable trim size, its endnotes are abbreviated to citations only. Readers who want both the portable print volume and the full annotations can consult this edition alongside it; the chapter and section structure is identical.

The book and both editions are also available online at:

`https://towardsliberty.com/pop`

Standards, tools, and case law referenced in the technical chapters change faster than any book can; the online page tracks updates and corrections.