

RANSOMWAVE

How Criminals Weaponise Credibility and
How Organisations Can Break Free



Security Digest Volume 1 | all rights reserved

INTRODUCTION



Setting the scene

A payments operations manager at a mid-sized bank arrives at her desk on a Tuesday morning and cannot log in. Neither can anyone else. The core banking platform is unreachable, transaction queues are frozen, and IT is getting the same error message on every terminal in the building. Within the hour, a message appears on a shared internal screen. The files are encrypted. A decryption key exists. And a figure, specific, not rounded, sits at the bottom of the note alongside a deadline and a set of instructions that are, against all expectation, clear and easy to follow.

This is not the chaotic, smash and grab scenario most people picture when they think about cybercrime. There is no visible perpetrator, no broken door, no obvious moment of intrusion to trace back to. There is just a frozen institution, a ticking clock, and a criminal organisation on the other end of an anonymised channel, waiting with the patience of a company that has done this before, because it has.

The bank's leadership now faces a decision that will be made under conditions designed to preclude good thinking: incomplete information, legal exposure, regulatory scrutiny, and the knowledge that every hour of downtime is measurable in lost transactions, customer complaints, and reputational damage that compounds. The ransom note, meanwhile, is professional. It may include a negotiation portal. In some cases, a live chat support function. The criminals are, on every surface measure, easier to deal with right now than the organisation's own recovery process.

That is the problem this paper addresses, not ransomware as a technical event, but ransomware as a trust operation. The encryption is the mechanism. The credibility is the weapon. Criminal groups have built a model that works not because it is technically unbreakable, but because it is psychologically coherent: victims pay because, under the conditions the attack deliberately creates, payment feels like the rational choice. Dismantling that logic, before the attack and not during it, is what the following pages are about.

Part one examines the criminal business model: the operational discipline, the pricing strategy, and the structural cracks that exist within it. Part Two turns to the victim's calculation, specifically why payment so often feels like the least bad option, even when the evidence is thin. Part Three goes deeper into the psychology of decision making under duress, tracing how fear, sunk cost, and institutional pressure combine to produce choices that later feel inexplicable. Part Four examines how cryptography itself becomes a credibility mechanism, and how the technical architecture of the attack is turned into the thing that makes victims trust their attackers. Part Five is the practical section: seven lessons for eliminating the need to trust criminals before the crisis begins.

The aim throughout is not to judge the organisations that pay. The conditions ransomware creates are engineered to overwhelm, and the people making these decisions are rarely reckless. They are experienced professionals who have run out of options at the worst possible moment. The real work is making sure those options exist in advance, so that when the note appears, the answer is already written.

Content

Part 1	The Criminal Business Model Behind the Ransom Note	06
Part 2	Why Victims Believe the Decryption Promise	13
Part 3	The Psychology of Trust Under Cyber Extortion	21
Part 4	How Cryptography Weaponises Credibility	32
Part 5	Eliminating the Need to Trust Criminals	39
	References	46
	Conclusion	51

\$1.1B

Global ransomware payments received in 2023 — up from \$567M in 2022.

Source: Chainalysis Crypto Crime Report 2024

2,825

Ransomware complaints filed with the FBI in 2023, with adjusted losses of \$59.6M in the US alone.

Source: FBI IC3 2023

85% → 29%

Victim payment rate, 2019 to Q4 2023. The single most powerful illustration of defensive posture change.

Source: Coveware Quarterly Reports

\$568,705

Average ransom payment in 2024. Median: \$200,000. The gap shows negotiation compression in action.

Source: Coveware 2024

Part I

The Criminal Business Model Behind the Ransom Note



Ransomware is routinely described as digital extortion; and that is accurate, as far as it goes. But the framing misses what makes it persist, scale, and survive law enforcement pressure decade after decade. Modern ransomware is not just malware. It is a market, with pricing strategies, customer management, brand reputation, service-level expectations, and multi-tier supply chains. And it functions, paradoxically, because victims and criminals share a fragile but essential ingredient: trust.

In a legitimate business, trust encourages customers to pay because they expect to receive the promised product. In ransomware, trust encourages victims to pay because they believe payment will restore access, stop further harm, and end the crisis. Criminals manufacture that trust deliberately, because without it, the business model collapses entirely.

THE INVISIBLE PHASE

1

1. Initial Access

Credential theft, phishing, or exploitation of unpatched internet-facing systems. The attack starts here — weeks or months before the victim is aware.

2

2. Reconnaissance

Attacker maps the environment: identifies valuable data, locates backup infrastructure, and tracks internal communications and escalation patterns.

3

3. Lateral Movement

Attacker moves through the network using legitimate credentials and tools to avoid detection. Access broadens progressively.

4

4. Backup Identification & Destruction

Backup servers are identified and destroyed or encrypted first. This is the move that eliminates the victim's primary alternative to payment.

"Every ransomware negotiation that ended in payment was preceded by a backup strategy that failed this test."

5

5. Data Exfiltration

Sensitive data is extracted to attacker-controlled infrastructure before encryption. This enables double and triple extortion.

ENCRYPTION TRIGGERED — THE VISIBLE PHASE BEGINS

6

6. Encryption Payload

Ransomware deployed across targeted systems. This is the visible, dramatic moment — but the attack was already complete before it arrived.

7

7. Ransom Note Delivered

Victim discovers the attack. Countdown timer begins. Negotiation portal opens.

8

8. Deliberation Window

72-hour average deliberation period. Boards, legal counsel, IT, insurance, and sometimes law enforcement are all involved.

9

9. Payment or Recovery

Victim pays and receives (or does not receive) a working decryptor — or activates independent recovery if backup infrastructure survived.

● Threat actor activity (invisible) ● Impact ○ Victim response

1.1

Why Trust Is the Engine, Not the Side Effect

At first glance, trust seems incompatible with crime. Victims are coerced, not persuaded. Yet ransomware operators face a genuine commercial problem: victims can choose not to pay. Payment is typically voluntary, even under extreme pressure, because an organisation may restore from backups, rebuild systems, involve law enforcement, or simply accept the downtime as the lesser cost.

For ransomware to remain profitable, attackers must persuade victims that paying is the rational option. That persuasion rests on credibility around three specific promises:

- **The decryption promise:** "Pay us, and you will receive a working decryptor and recover your files."
- **The non-disclosure promise:** "Pay us, and the stolen data will not be leaked or sold."
- **The non-revictimisation promise:** "Pay us, and we will not return immediately or keep extorting you."

If victims widely concluded these promises were false, payment rates would fall sharply and ransomware revenue would collapse. Ransomware groups therefore invest heavily in appearing reliable; not because they are ethical, but because credible commitment is profitable.

1.2

The Engineered Crisis: Encryption, Data Theft, and Manufactured Pressure

Ransomware pressure is not purely technical. It is a designed crisis, built to maximise the victim's perceived cost of not paying. Encryption converts operational capability into downtime. The product the attacker sells is not merely a decryptor; it is time, and time is money in every sector.

Many groups steal data before encrypting it, creating secondary pressure through reputational damage, regulatory exposure, and legal liability. Some go further, threatening to contact the victim's customers or partners directly; a third layer of coercion that is still trust-based, because the attacker must convince the victim they genuinely hold the data and genuinely intend to publish it if unpaid.

THE THREE EXTORTION LAYERS

- **Single extortion:** encrypt files, demand payment for the decryption key.
- **Double extortion:** steal data first, then threaten a public leak if unpaid.
- **Triple extortion:** contact the victim's own clients or partners directly.

Each layer adds leverage. Each requires trust to function.

- **Operators and affiliates:** deploy ransomware, manage lateral movement, and execute the compromise.
- **Developers:** maintain ransomware code, encryption tooling, and evasion capabilities.
- **Negotiators:** run victim-facing chat portals, manage bargaining, and provide what amounts to customer service.
- **Money movers:** launder proceeds through cryptocurrency mixing, exchanges, and mule networks.

This division of labour mirrors legitimate supply chains. Each role depends on reputation and reliability within the criminal ecosystem. A broker who sells bad access loses future business. An affiliate whose execution is sloppy gets cut off. A developer whose malware fails loses market share. Internal trust, criminal to criminal, is as structurally necessary as external trust; criminal to victim.

1.3

Ransomware Operates Like a Business; Because It Is One

Over time, ransomware groups have adopted business-like structures because specialisation increases revenue and reduces operational risk. The roles are distinct, repeatable, and recognisable to anyone familiar with how legitimate enterprises operate:

- **Initial access brokers:** obtain and sell network entry via stolen credentials, exposed VPN access, or prior compromise.

1.4

Ransomware-as-a-Service: Scaling Trust Across a Criminal Platform

Ransomware-as-a-Service (RaaS) is the clearest expression of ransomware as a business model. A core team provides the software and infrastructure; affiliates perform the intrusions and share the profits. This creates two distinct and interdependent trust relationships:

- **Affiliate-to-platform trust:** affiliates trust the RaaS operator will track payments honestly, provide functional tooling, and not steal their revenue share.
- **Victim-to-ecosystem trust:** victims trust that paying leads to decryption and a reduction in harm.

RaaS platforms often introduce features that mirror legitimate software-as-a-service products: dashboards, support channels, update notes, documentation, revenue splits, and operating rules for affiliate; including, in some cases, restrictions on targeting hospitals or critical infrastructure to avoid drawing unwanted law enforcement attention. This professionalisation makes the ecosystem more resilient, more scalable, and harder to dismantle.

1.5

Pricing: Optimisation, Not Guesswork

Ransom demands are rarely arbitrary. Many groups price based on the victim's revenue and ability to pay, the criticality of affected services, the estimated cost of downtime per day, the sensitivity of any exfiltrated data, known insurance coverage, and the perceived maturity of backups and recovery capability. Negotiation is built into the model. A demand set too high increases refusal. One set too low leaves revenue uncaptured. Skilled operators treat pricing like a sales optimisation problem: anchor high, negotiate down, close quickly, and compress the victim's perceived alternatives before they can think clearly.

1.6

Where the Model Breaks

Because the ransomware business model depends on credibility, it is also vulnerable to credibility failures. The following conditions weaken an attacker's ability to convert victims into payers:

- **Decryption unreliability:** broken decryptors or partial recovery damage a group's reputation across the ecosystem, reducing future conversion rates.
- **Post-payment betrayal:** data leaked despite payment, or a second extortion demand arriving shortly after, poisons the well for the entire ecosystem.

- **Fast victim recovery:** tested backups, segmented networks, and practised rebuild procedures reduce the attacker's leverage to near zero.
- **Sustained disruption pressure:** infrastructure takedowns, sanctions, arrests, and cryptocurrency wallet tracing raise operational costs and reduce reliability for affiliates.

Ransomware thrives when it can offer a believable bargain: "Pay, and this ends." Defensive strategy should aim to make that bargain less believable; or, better, to make it entirely unnecessary.

Part 2

Why Victims Believe the Decryption Promise

Ransomware only generates revenue if victims trust that paying will actually result in the return of their data. That trust does not arise from naivety or panic alone. It is the product of deliberate engineering; a combination of economic logic, psychological pressure, and carefully designed credibility signals. Understanding why victims believe the decryption promise is essential to understanding why ransomware remains so profitable, and what would need to change for it to become less so.

Many victims believe the decryption promise because the expected loss from not paying appears worse than the risk of being cheated. The calculation is not irrational. It is just being made inside a trap.

2.1

The Rational Calculation Under Impossible Conditions

Research into real-world ransomware incidents consistently shows that payment decisions are rarely impulsive. Boards, IT teams, legal counsel, insurance contacts, and sometimes law enforcement are all involved. Victims carefully analyse the situation before deciding – though they do so under time pressure, with incomplete information, and in conditions of acute stress that the attacker has deliberately created.

The calculation victims make is not irrational. They weigh the value of the encrypted data, the cost and time required to rebuild systems without a decryption key, the risk of regulatory penalties or reputational harm if stolen data is published, and the size of the ransom against the uncertain probability that decryption will actually work. When backups are missing, inadequate, or untested, this calculation tilts sharply toward payment. The expected loss from not paying appears worse than the risk of being deceived. That conclusion is not wrong; it is simply being reached in conditions the attacker deliberately engineered.

2.2

Loss Aversion, Framing, and the Psychology of the Ransom Note

Ransomware exploits one of the most well-documented principles in behavioural economics: loss aversion. People are significantly more motivated to avoid losses than to pursue equivalent gains. Victims facing concrete, immediate loss of critical data find payment psychologically attractive even when their rational assessment of the attacker's trustworthiness is low.

Experimental research on ransom splash screens; the messages victims see when they first discover the attack; demonstrates that how the demand is worded materially affects willingness to pay. Positively framed messages increase compliance. Urgency, professional language, and fear all nudge victims toward treating payment as the safest exit from a crisis. Timers and graphic elements matter less than the words themselves; attackers know this, and they write accordingly.

2.3

Criminal Reputation as a Trust Signal

From the attacker's perspective, honouring the decryption promise is often economically rational. Game-theoretic models of ransomware show that if victims expect decryption with high probability, more victims pay, which increases long-term revenue. In some market conditions, "always decrypt after payment" is the profit-maximising strategy.

A dedicated ransomware-and-reputation model identifies equilibria where the optimal strategy is to consistently return files, specifically to sustain belief in the promise. The ability to build this reputation depends on whether victims can distinguish between ransomware families and groups; through prior cases, leak sites, or advice from incident responders. That distinction becomes a trust signal in itself. Groups known to provide working decryptors can demand and receive higher ransoms, because victims believe paying will work.

Some groups therefore consistently decrypt not out of any ethical consideration, but to protect their most valuable business asset: a reputation for following through.

2.4

Information Asymmetry: The Victim Cannot Verify Anything

Victims operate under deep uncertainty. In double-extortion attacks, the attacker claims both to have encrypted critical files and to have exfiltrated sensitive data. Signalling-game analysis shows that criminals have strong incentives to claim exfiltration whether or not it actually occurred, because victims cannot verify the claim. The threat costs the attacker nothing to make and substantially increases leverage.

The same asymmetry applies to decryption. Victims cannot confirm in advance whether the attacker actually holds the keys, whether those keys will work on all affected systems, or whether full restoration is achievable. Multiple outcomes are possible, including ones where the attacker deceives from the start. Victims fill this information gap by assuming that "professional" groups will honour their commitments; a shortcut that the criminals deliberately cultivate through the signals described throughout this paper.

2.5

Criminal Reputation as a Trust Signal

From the attacker's perspective, honouring the decryption promise is often economically rational. Game-theoretic models of ransomware show that if victims expect decryption with high probability, more victims pay, which increases long-term revenue. In some market conditions, "always decrypt after payment" is the profit-maximising strategy.

A dedicated ransomware-and-reputation model identifies equilibria where the optimal strategy is to consistently return files, specifically to sustain belief in the promise. The ability to build this reputation depends on whether victims can distinguish between ransomware families and groups; through prior cases, leak sites, or advice from incident responders. That distinction becomes a trust signal in itself. Groups known to provide working decryptors can demand and receive higher ransoms, because victims believe paying will work.

Some groups therefore consistently decrypt not out of any ethical consideration, but to protect their most valuable business asset: a reputation for following through.

2.6

Information Asymmetry: The Victim Cannot Verify Anything

Victims operate under deep uncertainty. In double-extortion attacks, the attacker claims both to have encrypted critical files and to have exfiltrated sensitive data. Signalling-game analysis shows that criminals have strong incentives to claim exfiltration whether or not it actually occurred, because victims cannot verify the claim. The threat costs the attacker nothing to make and substantially increases leverage.

The same asymmetry applies to decryption. Victims cannot confirm in advance whether the attacker actually holds the keys, whether those keys will work on all affected systems, or whether full restoration is achievable. Multiple outcomes are possible, including ones where the attacker deceives from the start. Victims fill this information gap by assuming that "professional" groups will honour their commitments; a shortcut that the criminals deliberately cultivate through the signals described throughout this paper.

2.7

Social Proof: When Expert Advice and Market Signals Reinforce Payment

External actors can unintentionally strengthen a victim's belief in the decryption promise. Survey experiments among small and medium-sized enterprises show that being advised by a cybersecurity firm to pay significantly increases the stated likelihood of payment; the expert's endorsement functions as social proof.

Perceived past behaviour of specific ransomware groups also shapes expectations powerfully. Interviews with negotiators, law enforcement personnel, and victims reveal that knowledge of what a particular group has done before is among the strongest predictors of what a current victim decides to do. Bitcoin-tracing studies show that many campaigns receive substantial repeated payments over years — which implicitly signals to future victims that "others must be getting something for their money." Many failures, meanwhile, go unreported, systematically distorting the perceived success rate of payment.

2.8

The Limits of Belief: When the Promise Fails

Belief in the decryption promise is far from universal, and often misplaced. Technical analyses show that decryption reliability is highly family-specific: some strains can be reliably reversed using publicly available tools; others cannot, even with the key. Many victims who pay do not achieve full recovery.

Large-scale surveys show that many organisations still refuse to pay; particularly those with strong backup infrastructure, leadership that has pre-committed against payment, or ethical concerns about funding criminal activity. And studies of victim harm document severe operational, psychological, and in some cases physical consequences of ransomware incidents, confirming that these attacks are whole-of-organisation crises, not IT problems with a straightforward financial solution.

KEY RESEARCH FINDINGS

Payment decisions involve boards, legal, IT, insurance, and sometimes law enforcement; they are rarely impulsive.

Available, tested backups are the single strongest factor reducing likelihood of payment.

Positively framed ransom messages measurably increase victim compliance.

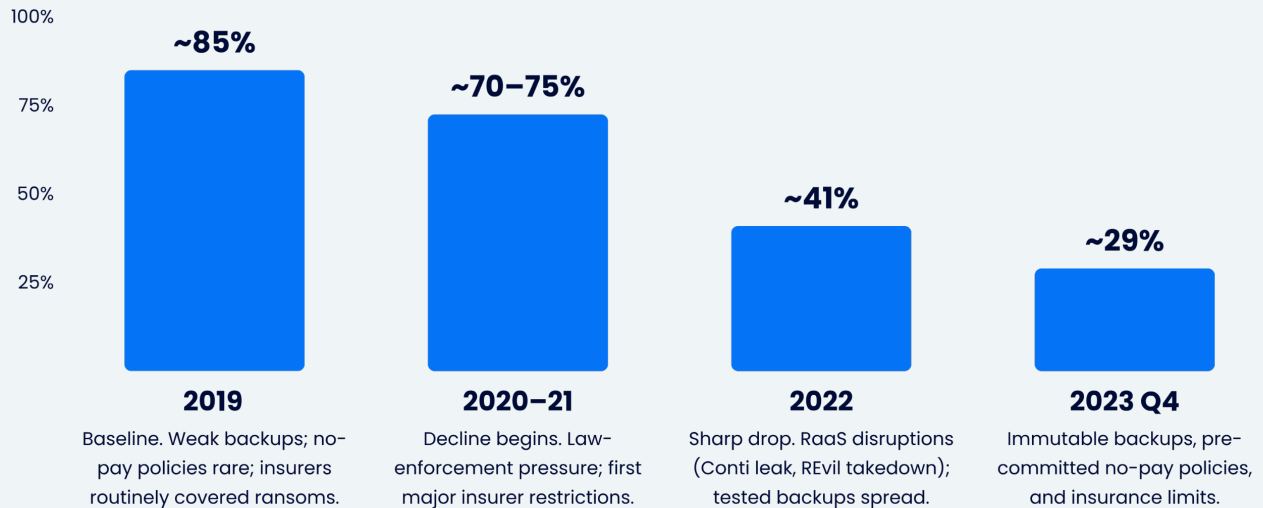
Cybersecurity firms advising payment significantly increase stated payment likelihood among SMEs.

Many decryption failures go unreported, skewing perceived success rates upward.

Ransomware causes severe operational, psychological, and physical harm — it is not just an IT issue.

PAYMENT RATES OVER TIME

Share of victims who paid the ransom (%)



SECTOR DISTRIBUTION OF ATTACKS

Relative targeting intensity — highest to lowest

Healthcare

Highest targeted — Operational continuity is life-critical, patient data highly sensitive, and security budgets historically under-resourced relative to exposure.

Education

Consistently top three — Decentralised IT, high user volumes, limited budgets, and dense repositories of personal data create a large attack surface.

Critical Infrastructure

Government, energy, water, transport — Regulatory consequences of downtime amplify payment pressure well beyond the ransom demand itself.

Financial Services

Increasingly targeted — Demands are calibrated to capacity to pay; exfiltration threats carry severe regulatory exposure under GDPR and equivalents.

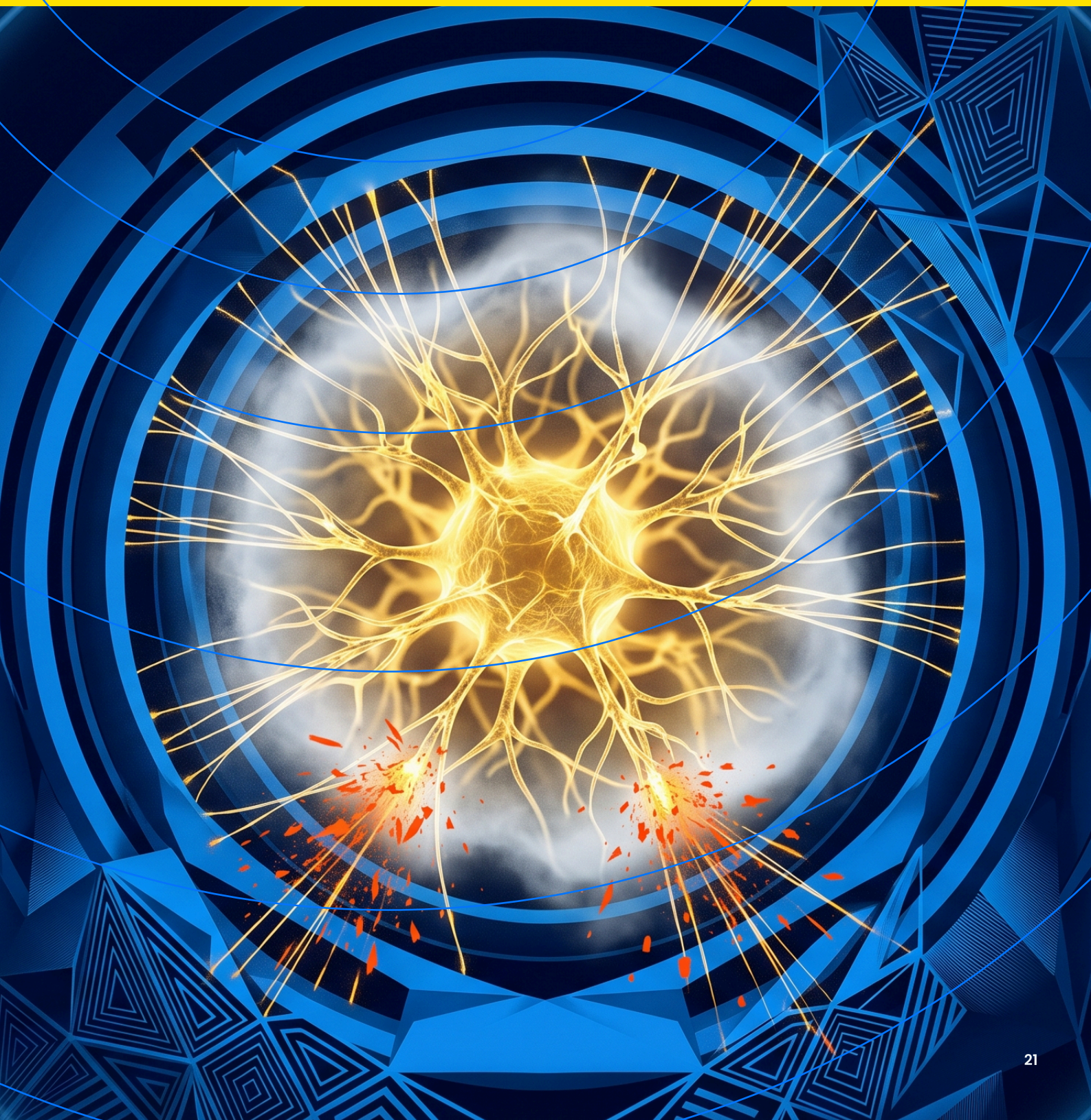
Manufacturing

Rising targeting — OT environments are frequently unpatched; downtime costs are immediate and quantifiable, which simplifies attacker pricing.

“Available, tested backups are the single strongest factor reducing likelihood of payment.”

Part 3

The Psychology of Trust Under Cyber Extortion



Cyber extortion forces victims into a psychological crisis unlike almost any other business emergency. They must decide whether to trust criminals who are actively harming them. Every rational indicator argues against it: the attackers are anonymous, hidden behind layers of technological obfuscation, operating entirely outside any legal framework that could enforce their promises. They have already demonstrated both the capability and the willingness to cause serious harm. They have every incentive to take payment and disappear.

Yet organisations routinely spend 72 hours, sometimes far longer, debating not whether to trust, but how much. This is not irrationality. It is trust formation operating under conditions where every conventional signal of trustworthiness has been stripped away. Understanding what fills that void explains a great deal about why ransomware keeps working.

1. The Trust Paradox: A Scenario That Should Not Exist

Consider a financial services firm that discovers three months of customer transaction data has been stolen. The attackers demand five million dollars. Pay, and the data will be deleted with proof of destruction. Refuse, and it will be sold to competitors and published on the dark web.

The firm's leadership knows the attackers are anonymous, unaccountable, and have already caused harm. They know there is no contract, no court, and no enforcement mechanism. They know the attacker has every incentive to take the money and violate every commitment. They know all of this; and still they spend three days debating terms. Not because they are naive. Because the alternative is psychologically untenable.

This is what makes cyber extortion structurally different from almost any other form of criminal activity. Traditional extortion; kidnapping, protection rackets, blackmail; involves ongoing relationships or physical proximity that create at least rudimentary alignment between criminal and victim interests. A kidnapper must keep the victim alive to collect. Cyber extortion eliminates these mechanisms entirely. The attacker has already obtained what they came for. Physical distance removes any retaliation risk.

Cryptocurrency makes payment irreversible and untraceable. The victim has no leverage except the attacker's voluntary adherence to their own stated terms.

2. The Cognitive Mechanics of Decision-Making Under Duress

2.1

Uncertainty Intolerance and the Need for Closure

Human minds are poorly equipped for sustained, unresolved uncertainty. Psychological research consistently shows that people experience ambiguity as more distressing than certain negative outcomes. Under cyber extortion, three scenarios face the victim:

Pay, and receive the decryption key. The crisis ends.

Pay, and be deceived. The money is gone. The situation may be worse.

Do not pay, and remain in an indefinite operational limbo with no clear end.

That third option is psychologically unbearable. The organisation cannot operate, cannot plan, cannot close the incident. Even the second option, which is financially worse, provides closure. The psychological relief of resolution, even negative resolution, exerts enormous pressure toward payment.

Paying becomes the only action that guarantees an end to uncertainty, regardless of whether that end is good.

2.2

The Illusion of Control Through Transaction

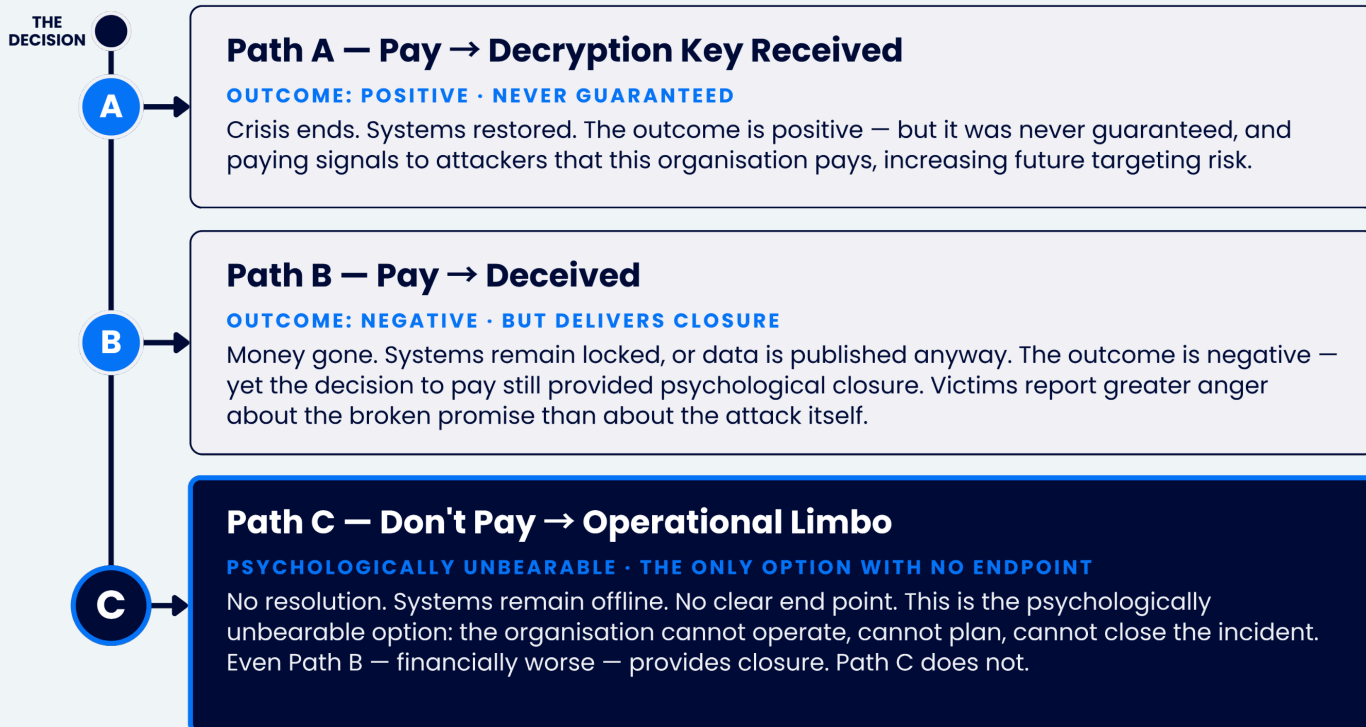
Payment is the one variable victims can control. They cannot control whether the attacker will honour their word, but they can control whether they initiate the transaction. Behavioural economists call this the preference for action: humans consistently prefer doing something over doing nothing, even when the action carries higher objective risk. The executive who authorises payment may not be trusting the attacker at all. They may be trusting their own decision-making; reclaiming agency in a situation that has stripped every other form of control away.

2.3

Sunk Cost and Escalating Commitment

Once negotiations begin, something shifts. Victims become psychologically invested in the belief that a resolution is possible; that the attacker is a rational, negotiable actor with whom a deal can be struck. Every exchange reinforces this: the attacker responds within a predictable timeframe, decrypts a small sample file as a show of good faith, answers technical questions coherently, and makes small concessions on price. None of this is evidence of trustworthiness. But it creates the appearance of a relationship, and human psychology automatically maps relationships onto trust frameworks. The longer negotiations run, the harder it becomes to walk away; not because evidence accumulates, but because cognitive investment does.

THE THREE DECISION PATHS



▲ The six forces below drive the victim out of Path C's limbo — toward paying (Path A or B). ▲

SIX PSYCHOLOGICAL FORCES

In order of activation — immediate onset (1) through to the negotiation phase (6)



1. Loss Aversion

We're wired to avoid losses far more than to chase equivalent gains.



2. Uncertainty Intolerance

Unresolved ambiguity feels worse than a certain bad outcome; paying ends it.



3. Illusion of Control

Payment is the one variable the victim controls — action over inaction.



4. Catastrophe Aversion

Facing total ruin, a slim recovery chance overrides expected-value logic.



5. Sunk Cost & Escalating Commitment

The longer negotiations run, the harder it becomes to walk away.



6. Diffusion of Responsibility

Shared accountability makes risky trust easier to authorise — 'we all agreed.'

"The calculation is not irrational. It is just being made inside a trap."

3. The Emotional Architecture of Adversarial Trust

3.1

Fear-Driven Trust Seeking

Victims of cyber extortion experience acute fear across multiple dimensions simultaneously: fear of permanent data loss, fear of reputational damage, fear of regulatory penalties, fear of operational collapse, and fear of personal accountability. This emotional state fundamentally alters decision-making. Under intense fear, humans engage in what psychologists call superstitious behaviour; taking actions that provide psychological comfort even when rational analysis suggests futility. Payment becomes a ritual act: "I have done everything in my power to save this situation."

This explains a pattern that puzzles many outside observers: victims paying despite explicit warnings from law enforcement, experienced incident responders, and their own internal advisors. The psychological need to act overwhelms the rational risk assessment. Fear is not a side effect of the attack. It is a core feature of the attacker's strategy.

3.2

Motivated Reasoning and the Hope-Trust Link

Hope and trust are not the same thing. But under cyber extortion, they become almost indistinguishable. Victims engage in motivated reasoning — the well-documented psychological tendency to believe information that aligns with desired outcomes. They want to believe that the attackers are professionals who maintain their reputation. They want to believe that recovery is possible. They want to believe that paying will resolve the crisis. These hopes become beliefs. These beliefs become trust. Not because evidence supports them, but because the alternative; accepting that recovery may be impossible regardless of what they do — is psychologically untenable.

3.3

Catastrophe Aversion and the Collapse of Normal Risk Assessment

Standard risk frameworks break down under ransomware. Normally, decision-makers weigh probability against impact. But when facing total organisational catastrophe, probability calculations become almost irrelevant.

If there is a 30% chance that payment results in decryption and a near-certain chance that non-payment results in permanent loss, the decision from a catastrophe-aversion perspective is payment. The victim is not trusting the attacker in any meaningful sense. They are accepting a substantial chance of additional loss in exchange for any non-zero chance of avoiding total catastrophe. This is existential gambling, and it is the rational response to the irrational situation the attacker has engineered.

4. How Attackers Manufacture Trustworthiness

If there is a 30% chance that payment results in decryption and a near-certain chance that non-payment results in permanent loss, the decision from a catastrophe-aversion perspective is payment. The victim is not trusting the attacker in any meaningful sense. They are accepting a substantial chance of additional loss in exchange for any non-zero chance of avoiding total catastrophe. This is existential gambling, and it is the rational response to the irrational situation the attacker has engineered.

THE TRUST-BUILDING TOOLKIT

- 1. Proof-of-decryption demos:** decrypting a small sample file at no cost to prove possession of the key.
- 2. Negotiation portals:** structured, ticket-style workflows that create the impression of legitimate process.
- 3. Professional communications:** calm, responsive, business-appropriate language that reduces victim panic.
- 4. Consistent behaviour:** countdown timers, staged threats, and public shaming that train future victims.
- 5. Social proof:** references to other victims who "recovered successfully" after paying.
- 6. Reciprocity pressure:** small good-faith gestures that create psychological obligation to reciprocate.

Victims frequently report that attacker professionalism increased their willingness to pay. One described their experience: "They had better customer service than our actual vendors. They responded faster, provided clearer instructions, and demonstrated more technical competence than our own IT team." That professionalism is strategic theatre; designed specifically to convert victim fear into victim trust by mapping a criminal interaction onto a familiar business transaction model.

5. The Organisational Dimension

Once humans extend trust, its violation causes emotional harm entirely independent of the material consequences. Victims consistently report greater anger about the broken promise than about the original attack itself.

5.1

Diffusion of Responsibility

In organisational contexts, ransom decisions rarely rest with a single person. Boards, executive teams, legal counsel, and cybersecurity experts all participate. This diffusion of responsibility creates its own dynamics. No individual bears full accountability for extending trust to criminals. The collective nature of the decision provides psychological cover: "We all agreed this was the best option available." Shared responsibility paradoxically makes risky trust decisions easier to authorise.

5.2

Legal and Fiduciary Pressure

Corporate leaders face legal obligations to protect shareholder value and stakeholder interests. If paying a ransom represents the path of least total harm, failure to pay could theoretically be characterised as a breach of fiduciary duty. This transforms the decision from an emotional calculation into a legal one; which allows leaders to authorise payment without psychologically framing it as trusting criminals. In their own internal narrative, they are executing their obligations to the organisation.

5.3

The Institutional Memory Problem

The aftermath of ransomware incidents creates a persistent distortion in organisational learning. When payment leads to successful recovery, organisations quietly celebrate the decision. When payment fails; when decryption keys do not work, or data is published anyway; organisations frequently conceal the outcome to avoid scrutiny. Successful payments are normalised; failed payments are buried. The result is a systematic skewing of what future decision-makers believe about whether paying works. The base rate of failure is structurally underreported.

Part 4

How Cryptography Weaponises Credibility



Parts One, Two, and Three explain why victims trust ransomware operators; the economic logic, psychological forces, and emotional dynamics that make payment feel rational even to intelligent people under extreme pressure. This part examines what makes the underlying threat feel inescapable in the first place. The answer lies in cryptography: the same mathematics designed to protect banking systems, government communications, and medical records, turned against the organisations it was built to serve.

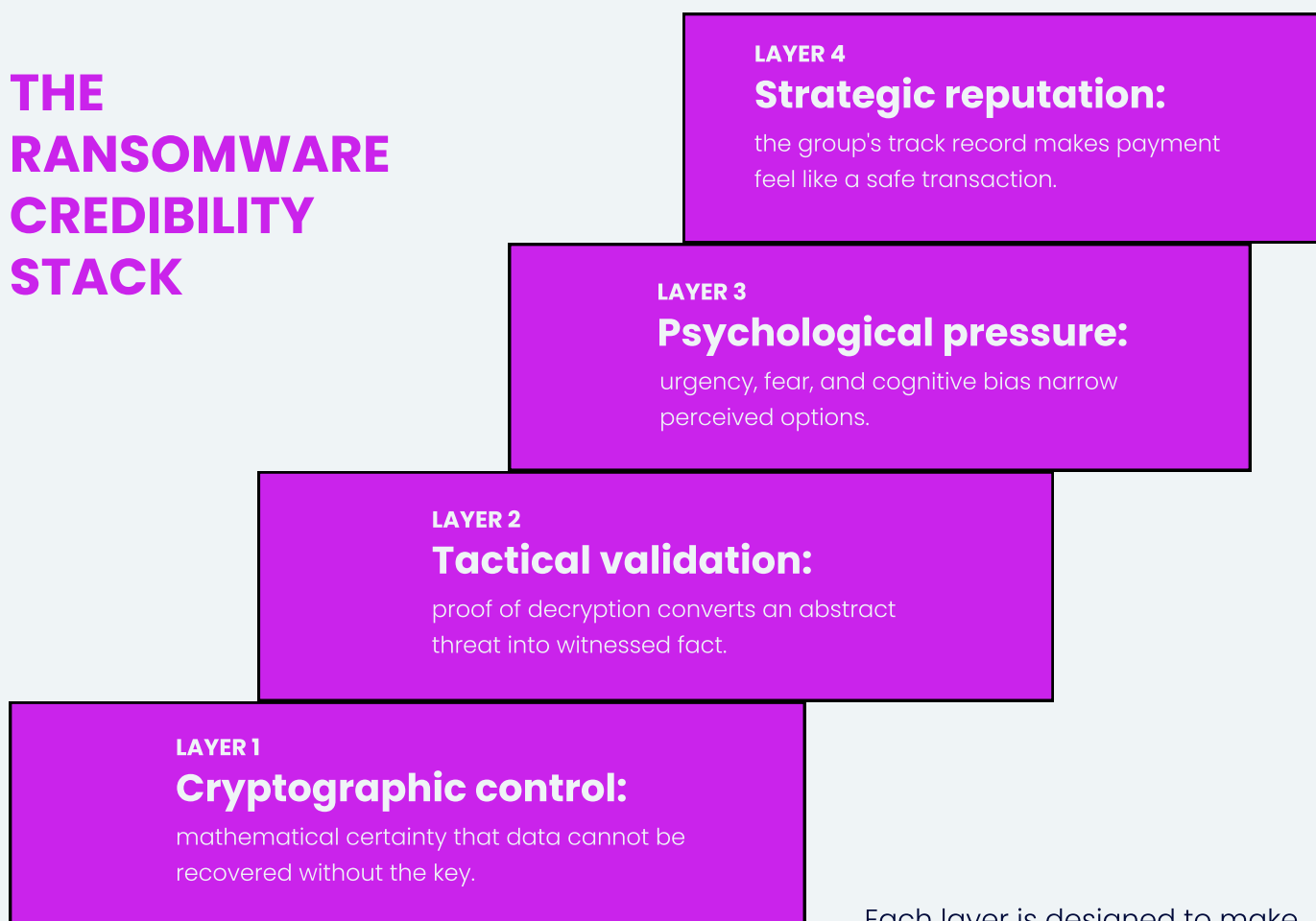
Ransomware's success is not purely an economic or psychological phenomenon. It rests on a technical foundation that produces something rare in the world of crime: mathematical certainty. When files are encrypted with algorithms like RSA-4096 or AES-256 — the same standards used by global financial institutions and intelligence agencies — the attacker is not merely locking data. They are invoking mathematical laws that no amount of computing power, technical skill, or determination can circumvent on any useful timescale. Brute-forcing a 256-bit AES key on a standard supercomputer would take longer than the current age of the universe. The question stops being "How do we fix this?" It becomes "What options do we have?" — a shift in framing that quietly moves every decision-maker closer to compliance before a word of negotiation has been exchanged.

4.1

The Credibility Stack: A Layered System of Coercion

Ransomware operators do not rely on cryptography alone. They construct a layered system of credibility in which each element reinforces the next, and the whole becomes far more coercive than any single component would be in isolation. Understanding this structure; what we can call the Ransomware Credibility Stack; is essential to understanding both how attacks succeed and where they are vulnerable.

THE RANSOMWARE CREDIBILITY STACK



Each layer is designed to make the layers above it more believable and more coercive.

At the base sits cryptographic control: the mathematical reality that recovery without the decryption key is not merely difficult, but practically impossible. This is the foundation on which everything else rests. Above it is tactical validation; the proof-of-decryption gesture that converts an abstract claim into a demonstrated fact. Above that is the psychological layer: urgency, fear, loss aversion, and artificial time pressure that discourage calm deliberation. At the top is strategic reputation; the accumulated history of a group that consistently decrypts after payment, making compliance feel like a rational business transaction rather than a desperate gamble.

4.2

Proof of Decryption: The Cryptographic Handshake

The most effective individual trust-building mechanism in ransomware is also the one most often underestimated by those who have not experienced an attack. When an attacker offers to decrypt a small set of files at no cost, they are not being generous. They are executing a calculated demonstration designed to eliminate the victim's most powerful source of doubt.

This act functions as a cryptographic handshake. It verifies, beyond any reasonable dispute, that the attacker holds the private key required for full recovery. And it does something more subtle: it anchors the victim's expectations.

What was previously uncertain — "Do they actually have the key? Will paying achieve anything?" — becomes demonstrably real. The system can be restored. The attacker holds the means to do so.

At this point, the decision to pay is reframed. It is no longer perceived as a gamble on an unknown outcome. It appears to be a transaction with a proven result. The psychological barrier to compliance drops sharply. Doubt is replaced with what might be called conditional trust — not trust in the attacker's character or intentions, but trust in the technical reality they have just demonstrated. This is precisely the shift the attacker is engineering.

4.3

The Professional Veneer: Customer Service as a Credibility Signal

Modern ransomware groups, particularly those operating under the RaaS model described in Part One, have adopted the surface features of legitimate corporate entities with deliberate precision. Victims are directed to polished web portals with structured support workflows. Negotiations are conducted in calm, formal business language. Deadlines are stated clearly. Extensions are offered. Concessions on price are made in a way that mirrors the give-and-take of a genuine commercial negotiation.



1. Proof-of-Decryption Demo

Attacker decrypts a small sample file at no cost. Converts an abstract claim — we hold your key — into a verified, witnessed fact. Eliminates the victim's most powerful source of doubt.

DOUBT ELIMINATED



2. Negotiation Portal

Victims are directed to a structured, ticket-style dark web portal with support workflows, response-time commitments, and a countdown timer. Creates the impression of a legitimate commercial process.

PROCESS LEGITIMISED



3. Professional Communications

Calm, responsive, business-appropriate language is used throughout. The attacker reduces victim panic deliberately — a panicked victim is harder to convert than a calculating one.

PANIC REDUCED



4. Consistent Behaviour

Countdown timers, staged threats, and public shaming of non-payers train future victims. The group's track record of following through on threats and promises is its most valuable asset.

CREDIBILITY PROVEN



5. Social Proof

References to other organisations that 'recovered successfully' after paying are introduced into negotiation. Functions as peer endorsement in a market with no other verification mechanism.

PEER-ENDORSED



6. Reciprocity Pressure

Small good-faith gestures — a decrypted file, a deadline extension, a price reduction — create psychological obligation to reciprocate. Human psychology maps these onto trust frameworks automatically.

OBLIGATION CREATED



PAYMENT AUTHORISED

THE ATTACKER'S OBJECTIVE

The victim trusts not the attacker's character, but the technical reality they have demonstrated.

None of this is coincidental. For a RaaS group, reputation is a revenue asset — possibly the most valuable one they possess. If a group fails to provide working decryption keys after payment, that failure propagates rapidly through the networks of incident response firms, cybersecurity forums, and the advisors who guide future victims. The group's conversion rate falls. Affiliates migrate to more reliable platforms. The brand is damaged in the only market that matters.

The profit-maximising strategy for a ransomware group is often to consistently return files after payment — to sustain belief in the promise. Cheating is only optimal in the very short term. It destroys long-run revenue. — Journal of Cybersecurity

This creates a paradox that victims find genuinely disorienting. The more professionally a criminal organisation behaves, the more likely victims are to trust it. One victim described the experience as receiving better customer service from their attackers than from their own software vendors — faster responses, clearer instructions, more apparent technical competence. That professionalism is strategic theatre. Its sole purpose is to convert victim fear into victim trust, by mapping a criminal interaction onto the familiar shape of a business transaction.

4.4

Disrupting the Stack: Breaking Cryptographic Credibility

Disrupting the Stack: Breaking Cryptographic Credibility

Understanding that ransomware credibility is a constructed system, not an inherent or inevitable feature of an attack, means it can be systematically dismantled. Each layer of the stack has a corresponding countermeasure, and each countermeasure, applied before an incident, reduces the coercive power of the layers above it.

- 1. Neutralise Layer 1:** cryptographic control. Immutable, offline backups are the only mechanism that makes the attacker's mathematical barrier irrelevant. If data can be restored independently, the impossibility of brute-force decryption becomes the attacker's problem, not the victim's.
- 2. Counter Layer 2:** tactical validation. Proof-of-decryption demonstrations should be understood as sales tactics, not evidence of trustworthiness. The willingness to decrypt three files does not predict the willingness — or even the technical ability — to decrypt thirty thousand.
- 3. Resist Layer 3:** psychological pressure. Leadership teams who have been trained to recognise urgency and scarcity as deliberate manipulation are significantly harder to move under artificial time pressure. Pre-incident decision frameworks, discussed in Part Five, are the primary countermeasure here.

4. Undermine Layer 4: strategic reputation. Sharing intelligence about attacker behaviour through threat intelligence platforms allows the security community to identify groups with poor post-payment track records, devaluing their most important business asset before the next victim encounters them.

The goal is not to neutralise cryptography; that is not achievable. The goal is to ensure that cryptographic control never reaches a victim who has no alternative path to recovery. Preparation breaks the stack at its foundation, before the upper layers can operate.

Part 5

Eliminating the Need to Trust Criminals



The question is not how to negotiate better with ransomware groups. It is not how to assess which groups are more likely to honour their commitments, or how to reduce the ransom to an acceptable figure. The right question is how to reach a position where the attacker's threats no longer matter; where the conversation inside the organisation shifts from "how much do we pay?" to "how quickly can we restore?"

That shift is entirely achievable. But it requires changing not just technology, but thinking. The lessons below draw on incident experience, organisational behaviour research, and the mechanics of the criminal business model described in the preceding parts. They apply to organisations of every size, across every sector. Some are technical. Some are human. All of them matter.

5.1

Lesson One: Accept That Criminals Lie

The most dangerous mistake a leader can make during a ransomware crisis is to treat the attacker like a business partner operating under reputational constraints. The reasoning is understandable: Parts One and Two of this paper explain why some groups do behave consistently. But consistency among criminals is a revenue strategy, not a moral commitment. It can change at any time, for any reason, with no warning and no recourse.

The digital extortion economy is lawless. When an organisation pays a ransom, it is not entering a contract. It is handing money to criminals and hoping. The extortion terms can shift mid-negotiation. Stolen data can be sold regardless of payment. A second demand can follow the first within weeks. None of this violates any agreement the criminal is legally obligated to keep, because no such obligation exists.

True organisational safety begins with a clear-eyed acceptance that paying a ransom guarantees nothing; and, by demonstrating that this organisation pays, it makes the organisation a more attractive target for subsequent attacks.

THE 3-2-1-1 BACKUP RULE

- 3 — Maintain three copies of all critical data.
- 2 — Store them on two different types of media (e.g., disk and tape).
- 1 — Keep one copy offsite.
- 1 — Keep one copy offline or in immutable storage, completely unreachable by attackers.

Every ransomware negotiation that ended in payment was preceded by a backup strategy that failed this test.

5.2

Lesson Three: Lock Down Data at the Source

Many organisations approach security like a perimeter defence: strong walls on the outside, open space within. Once attackers breach the perimeter; and with enough time and resources, they will; they move freely and identify the most valuable targets at their leisure. This model of security was inadequate a decade ago. Today it is indefensible.

The approach that counters it is called Zero Trust — a model built on the foundational assumption that attackers are already inside the network.

. Under Zero Trust, no user and no system is trusted automatically. Networks are divided into isolated segments. Users have access only to the specific files and systems their role requires. All sensitive data is encrypted, with the encryption keys held securely and separately from the data they protect.

The practical consequence: if an attacker manages to exfiltrate data, the files are unreadable. If they manage to deploy ransomware in one segment, it cannot spread freely to adjacent systems.

5.3

Lesson Three: Lock Down Data at the Source

Many organisations approach security like a perimeter defence: strong walls on the outside, open space within. Once attackers breach the perimeter; and with enough time and resources, they will; they move freely and identify the most valuable targets at their leisure. This model of security was inadequate a decade ago. Today it is indefensible.

The approach that counters it is called Zero Trust — a model built on the foundational assumption that attackers are already inside the network. Under Zero Trust, no user and no system is trusted automatically. Networks are divided into isolated segments. Users have access only to the specific files and systems their role requires. All sensitive data is encrypted, with the encryption keys held securely and separately from the data they protect.

The practical consequence: if an attacker manages to exfiltrate data, the files are unreadable. If they manage to deploy ransomware in one segment, it cannot spread freely to adjacent systems. Small breaches remain small. The mechanisms that allow a compromised workstation to become a total organisational catastrophe (lateral movement, broad access permissions, unencrypted sensitive data) are systematically removed.

5.4

Lesson Four: Practise Recovery Until It Is Routine

Technology alone does not protect organisations. People using technology, under pressure, in the dark, with phones ringing and executives demanding answers, protect organisations. When systems go dark and business halts, the quality of the response depends almost entirely on what the team has already rehearsed. A team that has never practised will panic. A team that has practised will execute.

A disaster recovery plan that exists as a document stored on a server is not a recovery plan. It is a document. The plan must be practised; repeatedly, with the actual people who will execute it under genuine stress. That means tabletop exercises and full simulation drills involving executives, legal teams, communications leads, and technical staff together. It means restoring data from backups in a controlled environment to verify the process works before it is needed. It means measuring the time that recovery actually takes against the organisation's stated recovery time objectives, and closing the gap.

There is also a psychological dimension to this. As Part Three explains, the cognitive pressure to pay is highest when a team has no other plausible path to recovery and no practised response to fall back on. Rehearsal removes that cognitive pressure.

The situation becomes a known problem with a known procedure; not an existential crisis where the criminal's payment portal appears to be the only exit.

5.5

Lesson Five: Control What the Attacker Knows

Sophisticated ransomware groups monitor their victims. Once inside a network, they read emails, observe internal communications, and track how much panic is building in leadership. They use this intelligence to calibrate their demands and time their pressure. An attacker who knows the board has convened an emergency session and the CEO has authorised a six-figure emergency budget has no reason to reduce the ransom. The victim's panic is visible, and it is being used against them.

Organisations must establish secure, out-of-band communication channels, methods of communicating that attackers cannot access even after compromising the primary network; before an incident occurs. The moment a breach is suspected, all sensitive recovery communications move to those channels. Recovery plans, decision-making discussions, forensic findings, and legal strategy are all conducted there, not on infrastructure the attacker may be reading in real time.

This deprives the attacker of their most valuable intelligence.

They are left guessing — about the organisation's resilience, its decision-making timeline, and its willingness to pay. That uncertainty is a meaningful defensive asset. Attackers who cannot read the room cannot calibrate their pressure effectively.

5.6

Lesson Six: Prepare Decision-Makers Before the Crisis Hits

The psychological forces described in Part Three — uncertainty intolerance, loss aversion, fear-driven trust, catastrophe aversion, and the illusion of control through transaction — do not disappear because an organisation has strong backups and good technology. They activate in human beings under pressure, regardless of what the technical posture looks like. The countermeasure is not better technology. It is better preparation of the humans who will make decisions.

Boards and executive teams should discuss and document their position on ransom payment before an incident occurs — when cognitive conditions are calm and deliberative, not during a crisis when every bias described in Part Three is operating simultaneously. That discussion should cover the circumstances under which payment might be authorised, the approval process required, who holds that authority, and what legal and insurance considerations apply. These questions have better answers when they are not being asked at two in the morning with the entire organisation offline.

Incident simulations should expose leaders to the psychological experience of a ransomware scenario; the countdown timers, the professional-sounding attacker communicating through a polished portal, the sense of powerlessness as systems remain dark; so that the experience is not entirely novel when a real attack occurs. Familiarity does not eliminate fear. But it substantially reduces the chance that fear drives a decision that calm deliberation would not have made.

5.7

Lesson Seven: Stop the Attack Before the Encryption Starts

The ransomware payload is the visible, dramatic part of an attack. But the attack itself began weeks or months earlier, when the attacker first obtained a foothold in the network. Stopping ransomware means stopping initial access; because once an attacker is inside, they have time to map the environment, identify valuable targets, disable defences, and destroy backups, all before deploying a single byte of encryption.

The most common initial access methods are credential theft, exploitation of unpatched internet-facing systems, and phishing.

The countermeasures are correspondingly straightforward: multi-factor authentication on all remote access, disciplined patch management that prioritises externally exposed systems, and user awareness training that focuses on recognising social engineering rather than generic cybersecurity hygiene. External exposure management, knowing precisely what your organisation looks like from the internet and systematically reducing that attack surface, is increasingly important as attackers use automated scanning tools to identify vulnerable targets at scale.

These controls are not glamorous. They do not appear in marketing materials or conference keynotes. But they are the ones that consistently determine whether an attacker reaches the point where the credibility stack, the proof-of-decryption demo, and the professional negotiation portal ever come into play.

CONCLUSION

The Bargain Becomes Irrelevant

Ransomware is not an unsolvable problem. It is a solvable one that most organisations approach too late, under the worst possible conditions, with the least possible clarity. That is not a technology failure. It is a preparation failure.

The analysis presented across the five parts of this paper converges on a single structural insight: ransomware works not because attackers are technically unstoppable, but because organisations allow themselves to arrive at the moment of crisis without a viable alternative to compliance. The criminal business model, the psychological architecture of extortion, the manufactured credibility of the decryption promise, the mathematical certainty of cryptographic control — all of these are powerful, but none of them are decisive in isolation. They become decisive only when combined with an unprepared victim.

The organisation that arrives at a ransomware incident with immutable backups tested within the last thirty days, a pre-agreed executive decision framework, segmented networks that prevented lateral movement, and a communications plan that operates outside the compromised environment does not face a crisis in the way the criminal business model assumes.

The attacker's toolkit — the countdown timer, the proof-of-decryption gesture, the polished negotiation portal, the carefully maintained reputation for following through — was built for an organisation that has no alternative. When an alternative exists, the entire credibility stack collapses. The bargain becomes irrelevant because there is no longer any need to accept it.

What This Paper Has Argued

Part One established that ransomware is a market, not merely an attack type. Criminal operators function within business structures that depend on reputation, division of labour, pricing optimisation, and trust maintenance. That trust is not incidental to the model. It is the engine of it. Without credibility, the decryption promise fails. Without a credible promise, payment rates collapse. Without payment, the enterprise dissolves. This means the model has structural vulnerabilities that organisations can exploit — but only if they understand it well enough to target them.

Part Two examined the victim side of that trust relationship. Payment decisions are rarely impulsive. They involve boards, legal counsel, insurers, and sometimes law enforcement. The calculation victims make is not irrational given the information they have at the time. What makes it dangerous is that the information they have at the time has been deliberately shaped by the attacker. The most important single finding from the research literature is also the most actionable: the availability of tested backups is the single strongest factor reducing the likelihood of payment. This is not a coincidence. It is the mechanism..

Part Three analysed the psychological forces that sustain trust under conditions where trust should be impossible.

Uncertainty intolerance, loss aversion, the illusion of control through transaction, sunk cost escalation, catastrophe aversion, motivated reasoning, and fear-driven decision-making all operate simultaneously during a ransomware incident. These forces do not disappear because an organisation has good technology. They activate in human beings regardless. The countermeasure is pre-incident preparation of the people who will make decisions under pressure — not better hardware, but better-rehearsed humans.

Part Four showed how cryptography transforms the threat from a claim into a demonstrated fact. The Ransomware Credibility Stack — cryptographic control, tactical validation, psychological pressure, and strategic reputation — is a constructed system, not a natural one. Each layer is designed to make the layers above it more coercive. But because it is constructed, it can be dismantled. Immutable backups neutralise the base layer. Pre-incident training inoculates against the psychological layer. Threat intelligence sharing degrades the reputational layer. The stack is not invulnerable. It is simply rarely attacked at the right point.

Part Five translated that analysis into seven organisational lessons: accepting that criminals lie, building backups that cannot be held hostage, locking down data at the source through Zero Trust architecture,

practising recovery until it is routine, controlling what the attacker knows through out-of-band communications, preparing decision-makers before the crisis hits, and stopping the attack before the encryption starts. None of these is a silver bullet. Collectively, they represent a coherent defensive posture designed to attack the criminal business model at its most vulnerable points.

The Stakes Have Changed

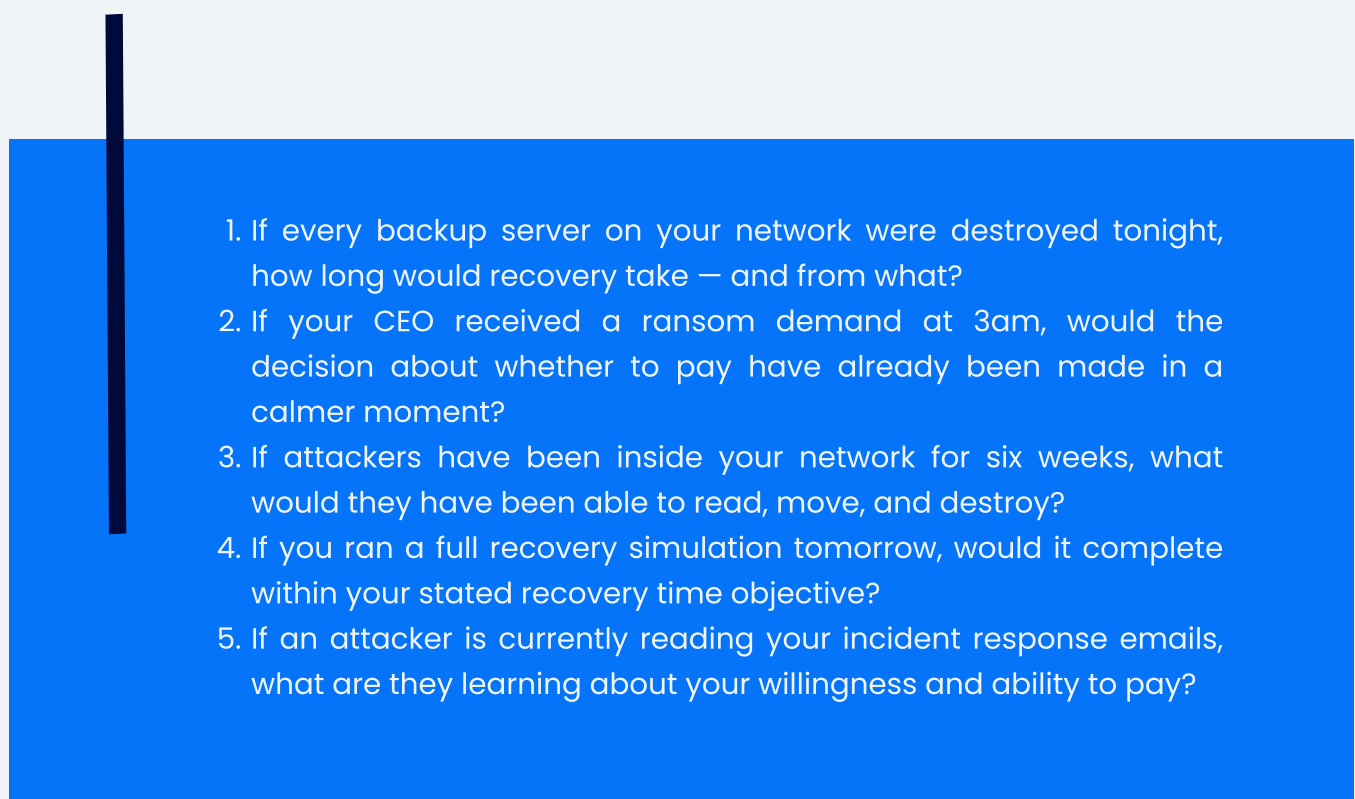
This paper was written at a moment when the ransomware threat has reached a degree of industrial maturity that demands a corresponding maturity in organisational response. The Ransomware-as-a-Service model has made sophisticated attacks accessible to affiliates who require no technical expertise of their own. Automation has made initial access scanning and exploitation faster than human defenders can respond manually. Artificial intelligence is beginning to be integrated into social engineering campaigns, making phishing more convincing and reconnaissance more targeted. The trajectory is clear.

At the same time, the defensive tools and frameworks available to organisations have never been more capable. Immutable cloud storage, identity-centric Zero Trust architectures, endpoint detection and response platforms with behavioural analytics, automated backup verification, and threat intelligence sharing platforms all represent meaningful advances over the defensive posture that was realistic even five years ago. The gap between what is possible and what most organisations have implemented remains the primary driver of ransomware's continued profitability.

"The goal is not to make ransomware impossible. It is to make the outcome of a ransomware attack boring: a known problem, with a practised response, that the organisation resolves independently without ever needing to consider whether a criminal's promise is worth trusting."

A Framework for the Decision Ahead

Every organisation that reads this paper faces a practical decision about where to invest attention and resource. The answer depends on where, along the chain of attack and response, the organisation is most exposed. The following framing is not a maturity model or a compliance checklist. It is a sequence of honest questions:

- 
1. If every backup server on your network were destroyed tonight, how long would recovery take – and from what?
 2. If your CEO received a ransom demand at 3am, would the decision about whether to pay have already been made in a calmer moment?
 3. If attackers have been inside your network for six weeks, what would they have been able to read, move, and destroy?
 4. If you ran a full recovery simulation tomorrow, would it complete within your stated recovery time objective?
 5. If an attacker is currently reading your incident response emails, what are they learning about your willingness and ability to pay?

These questions are not rhetorical. They have answers, and the answers determine how much of the criminal business model applies to your organisation. For organisations where the answers are satisfactory, the conclusion of this paper is a confirmation that the posture is correct. For organisations where the answers are uncomfortable, the conclusion is a call to act before the questions are asked under conditions the attacker has engineered.

The Responsibility of the Prepared

There is a dimension to this problem that extends beyond individual organisational risk. Ransomware is sustained, at a systemic level, by the aggregate payment behaviour of its victims. Every organisation that pays — even rationally, even under genuine duress — contributes to the credibility of the decryption promise for the next victim. Every payment validates the business model. Every successful recovery without payment degrades it.

This does not mean that organisations facing a genuine crisis, with no viable recovery alternative, are morally obligated to absorb catastrophic harm in the service of systemic deterrence. The individual calculus and the systemic calculus point in different directions in the short term. But the systemic point is important for organisations that are not yet in crisis: the investment in preparedness is not only a protection of the individual organisation. It is a contribution to the degradation of the criminal market.

Sharing incident intelligence through trusted channels, reporting attacks to law enforcement, disclosing post-payment failures rather than concealing them, and participating in threat intelligence platforms all contribute to a shared defensive commons. The criminal ecosystem depends on information asymmetry.

Defenders who reduce that asymmetry — even incrementally — shift the conditions under which the next attack will operate.

Final Assessment

Ransomware endures not because it cannot be stopped, but because the conditions that make it profitable have been allowed to persist. Those conditions — unprepared victims, untested backups, unpractised decision-makers, uncontrolled attacker access to organisational intelligence — are not natural phenomena. They are the result of choices that can be reversed.

The criminals who operate in this space are sophisticated, adaptive, and highly motivated. They invest in credibility the way legitimate businesses invest in brand. They understand their victims' psychology better than most organisations understand their own. They have optimised a business model that has generated billions in proceeds across more than a decade of operation.

But the model has a weakness. It requires victims who have no other option. The entire architecture of manufactured trust — the professional portals, the proof-of-decryption gestures, the carefully cultivated reputation for following through — exists because the attacker needs the victim to cooperate. That need is the vulnerability.

Organisations that build genuine independence from the decryption key do not need to assess whether the criminal is trustworthy. They do not need to calculate whether the ransom is worth paying. .

They do not need to spend seventy-two hours debating terms with someone who has no legal obligation to honour any of them. They restore from backup, contain the incident, identify the initial access vector, and move on. The criminal's carefully constructed credibility was never tested, because it was never needed.

That is the goal. Not to defeat ransomware in some abstract sense. Not to make attacks less likely — though good hygiene does that too. But to reach a position where, if an attack succeeds, the organisation already has everything it needs to recover. Where the attacker holds an encrypted copy of data the organisation can restore independently. Where the bargain being offered is one the organisation has already made unnecessary.

ORGANISATIONAL READINESS

THE CORE QUESTIONS

- 1. Recovery independence:** Can you restore critical systems from backups the attacker cannot reach?
- 2. Decision readiness:** Has your leadership pre-agreed its position on payment before a crisis hits?
- 3. Access containment:** Would a compromised endpoint become a total network compromise?
- 4. Recovery rehearsal:** Has your team actually practised restoring systems under simulated pressure?
- 5. Intelligence security:** Are your incident response communications on channels the attacker cannot read?
- 6. Awareness investment:** Do your staff recognise the social engineering techniques that create initial access?
- 7. Pre-encryption detection:** Do you have the detection capability to catch an attacker before encryption begins?

"When the organisation can answer yes to those questions, ransomware becomes what it should always have been: a known threat, with a known response, and a known outcome that does not involve trusting criminals."

"When the organisation can answer yes to those questions, ransomware becomes what it should always have been: a known threat, with a known response, and a known outcome that does not involve trusting criminals."

Infoassure works with organisations across sectors to assess, design, and implement the technical and human defences described in this paper. For enquiries, engagements, or to discuss the findings in the context of your organisation, visit **infoassureltd.com**.

REFERENCES

Academic & Peer-Reviewed Sources

- Cartwright, A., & Cartwright, E. (2019). Ransomware and reputation. *Games*, 10(2), 26. <https://doi.org/10.3390/g10020026>
- Cartwright, E., Castro, J., & Cartwright, A. (2019). To pay or not: game theoretic models of ransomware. *Journal of Cybersecurity*, 5(1), tyz009. <https://doi.org/10.1093/cybsec/tyz009>
- Cartwright, A., Cartwright, E., Xue, L., & Hernandez-Castro, J. (2022). An investigation of individual willingness to pay ransomware. *Journal of Financial Crime*. <https://doi.org/10.1108/jfc-02-2022-0055>
- Cartwright, E., Cartwright, A., & Xue, L. (2025). Ransomware splash screens, loss aversion and trust: insights from behavioral economics. *Journal of Cybersecurity Privacy*, 5, 69. <https://doi.org/10.3390/jcp5030069>
- Connolly, A., & Borrion, H. (2022). Reducing ransomware crime: analysis of victims' payment decisions. *Computers & Security*, 119, 102760. <https://doi.org/10.1016/j.cose.2022.102760>
- Connolly, L., & Wall, D. (2019). The rise of crypto-ransomware in a changing cybercrime landscape: taxonomising countermeasures. *Computers & Security*, 87. <https://doi.org/10.1016/j.cose.2019.101568>
- Conti, M., Gangwal, A., & Ruj, S. (2018). On the economic significance of ransomware campaigns: a Bitcoin transactions perspective. *ArXiv*, abs/1804.01341. <https://doi.org/10.1016/j.cose.2018.08.008>
- Delgado-Mohatar, O., Camara, J., & Anguiano, E. (2020). Blockchain-based semi-autonomous ransomware. *Future Generation Computer Systems*, 112, 589–603. <https://doi.org/10.1016/j.future.2020.02.037>
- Dey, D., & Lahiri, A. (2025). "Extortionality" in ransomware attacks: a microeconomic study of extortion and externality. *Information Systems Research*. <https://doi.org/10.1287/isre.2024.1160>
- Hernandez-Castro, J., Cartwright, A., & Cartwright, E. (2020). An economic analysis of ransomware and its welfare consequences. *Royal Society Open Science*, 7. <https://doi.org/10.1098/rsos.190023>
- Matthijse, S., Moneva, A., Van 't Hoff-De Goede, M., & Leukfeldt, E. (2024). Examining ransomware payment decision-making among small- and medium-sized enterprises. *European Journal of Criminology*, 22, 625–645. <https://doi.org/10.1177/14773708241285671>
- McIntyre, D., & Frank, R. (2021). No gambles with information security: the victim psychology of a ransomware attack. In *Cybercrime in Context* (pp. 43–60). Springer. https://doi.org/10.1007/978-3-030-60527-8_4
- Meurs, T., Cartwright, E., Cartwright, A., Junger, M., & Abhishta, A. (2023). Deception in double extortion ransomware attacks: an analysis of profitability and credibility. *Computers & Security*, 138, 103670. <https://doi.org/10.1016/j.cose.2023.103670>

- Mott, G., Turner, S., Nurse, J., Pattnaik, N., MacColl, J., Huesch, P., & Sullivan, J. (2024). Ransomware harms and the factors that influence the victim organization's experience. *Journal of Cybersecurity*, 10. <https://doi.org/10.1093/cybsec/tyae013>
- Moussaileb, R., Cuppens-Boulahia, N., Lanet, J., & Boudier, H. (2021). A survey on Windows-based ransomware taxonomy and detection mechanisms. *ACM Computing Surveys*, 54, 1–36. <https://doi.org/10.1145/3453153>
- Paquet-Clouston, M., Haslhofer, B., & Dupont, B. (2018). Ransomware payments in the Bitcoin ecosystem. *ArXiv*, abs/1804.04080. <https://doi.org/10.1093/cybsec/tyz003>
- Vakiliinia, I., Khalili, M., & Li, M. (2021). A mechanism design approach to solve ransomware dilemmas. In *GameSec 2021* (pp. 181–194). Springer. https://doi.org/10.1007/978-3-030-90370-1_10
- Yilmaz, Y., Cetin, O., Arief, B., & Castro, J. (2021). Investigating the impact of ransomware splash screens. *Journal of Information Security and Applications*, 61, 102934. <https://doi.org/10.1016/j.jisa.2021.102934>
- Yilmaz, Y., Cetin, O., Grigore, C., Arief, B., & Hernandez-Castro, J. (2022). Personality types and ransomware victimisation. *Digital Threats: Research and Practice*, 4, 1–25. <https://doi.org/10.1145/3568994>
- Yuste, J., & Pastrana, S. (2021). Avaddon ransomware: an in-depth analysis and decryption of infected systems. *ArXiv*, abs/2102.04796. <https://doi.org/10.1016/j.cose.2021.102388>

Policy & Standards Sources

- Cybersecurity and Infrastructure Security Agency. (2023). Stop ransomware guide. U.S. Department of Homeland Security.
- European Union Agency for Cybersecurity. (2023). ENISA threat landscape report. ENISA Publications.
- National Institute of Standards and Technology. (2020). Zero trust architecture (SP 800-207). U.S. Department of Commerce.
- World Economic Forum. (2024). Global cybersecurity outlook. WEF Publishing.

Industry Sources

- Cyber Threat Alliance. (2026). Ransomware threat intelligence sharing report. <https://www.cyberthreatalliance.org>
- DestCert. (2026). Asymmetric cryptography: RSA, ECC & PKI explained. <https://destcert.com/resources/asy>
- Fortinet. (2025). What is ransomware-as-a-service (RaaS)? How does it work? <https://www.fortinet.com/resources/cyberglossary/ransomware-as-a-service-raas>
- Palo Alto Networks Unit 42. (2026). Inside ransomware negotiations: trust criminals or walk away? [Threat Vector Podcast]. <https://www.paloaltonetworks.com/resources/podcasts/threat-vector-inside-ransomware-negotiations-trust-criminals-or-walk-away>