OMXUS PRESS • PREPRINT | Research article

Mathematical Foundations of Decentralised Trust: From Bitcoin's Proven Security to Social Consensus

A. C. Applebee[a] and L. N. Combe[a]

Preprint, not peer reviewed • Version 1.0-preprint • 2026-10-14 • Licence CC BY 4.0

Abstract

Trust without an authority can be built in two layers: a computational layer whose security is proven, and a social layer of in-person vouching whose security rests on assumptions about people. This paper states every theorem and claim the two layers need, marks each as proven, bounded under stated assumptions, heuristic, conjectured or open, proves what can be proved, and checks the rest with reproducible numerical models (none of them empirical data). The Bitcoin results hold: the exponential waiting-time law is within π of exact for a per-hash success chance π; the exact double-spend formula agrees with 100,000-race simulations; selfish mining never beats honest publication without difficulty adjustment (0 of 21 cells) and pays above the published threshold with it (14 of 14). A record anchored in Bitcoin inherits that bound: 12 confirmations hold a 10% attacker under one in a million. On the social layer, n fakes need at least nk/m real accomplices; collusion of three independent vouchers runs at 10^{-6}, rising about 200-fold under modest correlation; a rule that vouchers may not know each other leaves a fully knit ring no admissible vouch. Two results are new. A ripple of consequences sent up the sponsor chain at one third per sponsor carries at most one unit per generation, so total liability stays at most five at the design's depth of four, at any network size. And clustering provably lowers responder coverage at a fixed place below the Poisson value, while coverage at the person in trouble stays close to it.

Keywords: decentralised trust | Bitcoin | double spend | Sybil resistance | web of trust | vouching | ripple responsibility | Poisson processes | Cox processes | security bounds

Significance. People are asked to trust money, votes and identity systems on the strength of claims nobody separates into what is proven and what is hoped. This paper does that separation for one design: a Bitcoin-anchored record and an identity layer built on in-person vouching. It proves what can be proved, shows the arithmetic for what is only bounded, and leaves the open problems open. Two of the draft's open questions turn out to have answers: the vouching ripple is bounded at any scale, and clustering lowers coverage of empty places but not of the people who need help.

Introduction: two kinds of proof

Bitcoin's white paper ends with a table (Nakamoto, 2008). Under the heading *Calculations*, a pseudonymous author worked out how likely an attacker holding a tenth of the network's computing power is to rewrite a payment after five confirmations, and printed the answer to seven decimal places: 0.0009137. He printed the C code too. Anyone could check it.

Ten years later two mathematicians did. Grunspan and Pérez-Marco (2018a) solved the attacker's race exactly and showed that Nakamoto's approximation is a little optimistic for large attackers; the exact answer is a regularised incomplete beta function. The point is not the correction. The point is that a correction was possible at all. Bitcoin's security is the kind of claim that can be proven, measured in hash operations, and checked by a stranger with a laptop.

Social trust is not like that. When a system asks three people to vouch, in person, that a fourth is a real and single human being, its security rests on what those people will and will not do. Nobody can write that down as a hardness assumption. Douceur (2002) showed that without some cost tied to identity, one adversary can pose as as many identities as they like; every defence since has had to say what that cost is and why it holds (Viswanath et al., 2010; Yu et al., 2006; Yu et al., 2008). Classical agreement protocols assume the opposite problem away: they count on a known set of participants, a minority of whom may lie (Lamport et al., 1982).

This paper takes a design with both layers (a social identity layer built on in-person vouching, whose records are anchored in Bitcoin) and does one thing: it separates what is proven from what is hoped. Each result carries its status.

- **Theorem**: proven, with a complete proof here or in the cited literature.
- **Proposition, Lemma, Corollary**: supporting results with proofs.
- **Security Bound**: an inequality that holds under explicitly stated assumptions.
- **Heuristic**: a plausible argument without a formal proof.
- **Conjecture**: an unproven claim we believe to be true.
- **Open Question**: a problem requiring further research.

Every proven result is also checked numerically, and every bound is stress-tested by changing the assumption it rests on. The four studies are verification of mathematics: models and simulations, labelled as such, never empirical data. They are in the paper's repository with the code that produces every number. The statistics appendix sets out each check and the sceptic's questions.

Table 1 is the ledger. It lists every claim the earlier version (2.0, February 2026) and its claims register made, with the status each reaches here.

Table 1. Every claim, with its status in version 2.0 and here. Version 2.0's claims register is in the repository (`data/proof_status_source.csv`).

Claim	Version 2.0	This version	Where
Mining times exponential, block counts Poisson	Theorem	Theorem; error bound added (Proposition A)	Mining model, Study 1
Double-spend probability $I_{4pq}(z, 1/2)$	Theorem	Theorem; verified by simulation	Double spend, Study 1
Honest publication optimal without difficulty adjustment	Theorem	Theorem; verified by simulation	Stability, Study 1
Anchored record inherits Bitcoin's bound	Theorem	Theorem	Anchoring, Study 1
Attack cost bounded by mining difficulty; anchor order preserved	Proven	Theorem (follow from the above)	Anchoring
Sybil cost linear in fakes	Security bound	Security bound; selection sensitivity	Web of trust, Study 2
Fraud $\leq p_c^k$ under independence	Security bound	Security bound; correlation sensitivity	Verification, Study 2
Non-adjacent vouchers raise collusion cost	Heuristic	Heuristic; exact cost for random groups (Proposition 4)	Verification, Study 2
Ripple penalty bounded	Proposition	Proposition; sharpened; up-chain form (Proposition 3, new)	Ripple, Study 3
Poisson responder coverage	Proposition	Proposition	Emergency response, Study 4
Clustered coverage below Poisson	Conjecture	Theorem 6 (proved here)	Emergency response, Study 4
Urban coverage achievable at 1-5% adoption	Conjecture	Conjecture; supported by model	Emergency response, Study 4
BLS and Ed25519 signatures secure; content addressing sound	Proven	Proven in the literature under standard assumptions	Anchoring
Zero-knowledge proofs scale	Conjectured	Conjectured	Open problems
Privacy against verifiability; long-run incentives; cross-cultural trust	Open	Open	Open problems

Part I. The computational layer

The results in this part are rigorous, building on Nakamoto (2008), Rosenfeld (2014) and Grunspan and Pérez-Marco (2018a, 2018b, 2020). Broader security analyses of the protocol, the backbone properties of common prefix and chain quality (Garay et al., 2015) and their extension to networks with delay (Pass et al., 2017), rest on the same probability; the results here are restated so that the social layer can be measured against them.

The mining model

A miner holds a share $0 < p \leq 1$ of the network's hashrate. The network finds a block on average every $\tau_0 = 10$ minutes.

Theorem 1 (mining time). The time T a miner with share p waits for a block is exponential, $f_T(t) = \alpha e^{-\alpha t}$ with $\alpha = p/\tau_0$.

Proof. Each hash attempt is an independent trial that succeeds with probability $\pi = \text{target}/2^{256}$, because SHA-256 output behaves as uniformly random. A miner doing h hashes a second waits a geometric number of attempts; in the continuous limit the waiting time is exponential with rate $\alpha = h\pi$, and it is memoryless because the attempts are independent. □

The continuous limit can be made exact. The following bound is ours.

Proposition A (how close the limit is). Let the number of attempts N be geometric with success probability $\pi \leq 1/2$, and $T = N/h$. Then for every $t \geq 0$,

$$\left|\Pr(T > t) - e^{-h\pi t}\right| \leq \pi.$$

Proof. Write $n = \lfloor th \rfloor$. Then $\Pr(T > t) = (1-\pi)^n$, and $e^{-h\pi t}$ lies in $(e^{-\pi(n+1)}, e^{-\pi n}]$. Since $\ln(1-\pi) \geq -\pi - \pi^2$ for $\pi \leq 1/2$, $(1-\pi)^n \geq e^{-\pi n} e^{-\pi^2 n}$, so $e^{-\pi n} - (1-\pi)^n \leq e^{-\pi n}(1 - e^{-\pi^2 n}) \leq \pi^2 n e^{-\pi n} \leq \pi/e$. In the other direction, $(1-\pi)^n - e^{-\pi(n+1)} \leq e^{-\pi n}(1 - e^{-\pi}) \leq \pi$. □

At Bitcoin's lowest possible difficulty one hash in about 2^{32}, roughly four billion, meets the target, and difficulty has only risen since. The exponential law is therefore exact to better than one part in four billion. Study 1 confirms the bound is tight: the largest gap equals π to four significant figures at every π from 10^{-6} to 0.1.

Corollary 1 (block counts). The number of blocks $N(t)$ found by time t is Poisson: $\Pr[N(t) = n] = (\alpha t)^n e^{-\alpha t}/n!$.

In 200,000 simulated hours the mean was 6.005 blocks an hour, the variance 5.998 (dispersion 0.999) and a chi-square test against Poisson(6) gave $p = .85$ (Study 1). On real blocks the same law holds: the related paper *The Immutable Chain* tests it on all 2,016 blocks of the first difficulty epoch after the April 2024 halving.

The double-spend problem

An attacker with share $q < 1/2$ (honest share $p = 1 - q$) secretly mines a chain in which a payment never happened, and tries to overtake the honest chain after the merchant has seen z confirmations.

Theorem 2 (double-spend probability; Grunspan & Pérez-Marco, 2018). The attacker succeeds with probability

$$P(z) = I_{4pq}\left(z, \tfrac{1}{2}\right), \qquad I_x(a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x t^{a-1}(1-t)^{b-1}\, dt.$$

Proof. While the honest chain finds z blocks the attacker finds m blocks, a negative binomial count with parameters (z, p). From a deficit of $z - m$ blocks, the attacker catches up with probability $(q/p)^{z-m}$ (the gambler's ruin), or at once if $m \geq z$. Summing over m gives a finite series, which equals the regularised incomplete beta function; the full proof is in Grunspan and Pérez-Marco (2018a). □

Study 1 checks the identity between the closed form and the finite sum on a grid of 1,440 cells (q from 0.02 to 0.48, z from 1 to 60): the largest difference is 10^{-14}, rounding error. It then simulates the race with no formula at all, 100,000 races per cell, and the simulated rate sits within 1.2 standard errors of the formula in all five test cells.

Corollary 2 (exponential security). Let $s = 4pq < 1$. As $z \to \infty$,

$$P(z) \sim \frac{s^z}{\sqrt{\pi(1-s)z}}.$$

Proof. Watson's lemma applied to the integral form of $I_x(a,b)$ (Grunspan & Pérez-Marco, 2018a). □

The approximation is a limit, and the approach is slow for large attackers. The ratio of exact to asymptotic is 0.94 at six confirmations for a 10% attacker but only 0.55 for a 40% attacker, and still 0.96 for the 40% attacker at $z = 300$ (Study 1). The shape is the promise: for every attacker below half, risk falls by a factor of about $4pq$ with each block.

Table 2. Double-spend probability $P(z) = I_{4pq}(z, 1/2)$ for attackers holding 10% to 40% of the hashrate. Computed in Study 1 and checked against simulated races.

z	$q = 0.10$	$q = 0.20$	$q = 0.30$	$q = 0.40$
1	0.2000	0.4000	0.6000	0.8000
2	0.0560	0.2080	0.4320	0.7040
3	0.0171	0.1158	0.3262	0.6349
4	0.0055	0.0667	0.2521	0.5796
5	0.0018	0.0392	0.1976	0.5331
6	0.0006	0.0233	0.1564	0.4930

Version 2.0 printed a table under this theorem with smaller values (at six confirmations, 0.0003 for $q = 0.10$ and 0.0308 for $q = 0.30$). Study 1 sets it beside the exact values and tries six codings to reproduce it, including Nakamoto's approximation, the gambler's ruin alone, the asymptotic form and an offset in z. None reproduces all 24 cells. At $q = 0.10$ the draft sits on Nakamoto's approximation (0.2045 against 0.2046 at $z = 1$), so it is likely an approximation-based coding; its exact basis is not traced. The theorem itself is unchanged and the paper reports the exact values; the draft table is kept in the repository as a sensitivity record. Probability is not the whole price, either: counting the block rewards an attacker gives up, double-spending a modest payment does not pay for a small miner (Grunspan & Pérez-Marco, 2022). The conclusion the draft table was drawn to support stands in a sharper form: six confirmations are overwhelming protection against a small attacker, and the rate of protection is set by how concentrated mining power is.

Protocol stability

Theorem 3 (optimal strategy; Grunspan & Pérez-Marco, 2018). Without difficulty adjustment, a miner's best strategy is to publish every block the moment it is found.

Proof. Measure a strategy by its revenue ratio $\Gamma = \mathbb{E}[R]/\mathbb{E}[\tau]$ over a cycle of duration τ, a stopping time. $M(t) = N(t) - \alpha t$ is a martingale, so Doob's optional stopping theorem (Williams, 1991) gives $\mathbb{E}[N(\tau)] = \alpha\,\mathbb{E}[\tau]$: no strategy can find blocks faster than the miner's hashrate allows. Withholding can only lose blocks to orphaning, so $\Gamma \le pb/\tau_0 = \Gamma_H$ for block reward b, with equality only for immediate publication (Grunspan & Pérez-Marco, 2018b). □

Remark (the door difficulty adjustment opens). Difficulty adjustment retimes the chain to one block every ten minutes whatever is orphaned. A pool that withholds blocks to orphan honest ones then earns a larger *share* of a chain that still pays out at the same rate. Eyal and Sirer (2018) showed selfish mining pays once the pool's share exceeds $(1-\gamma)/(3-2\gamma)$, where γ is the fraction of honest miners who build on the pool's block in a tie: one third at $\gamma = 0$, one quarter at $\gamma = 1/2$. Better withholding strategies lower the threshold further (Sapirshtein et al., 2017). The gap between the theorem and the deployed protocol is real, and it is about power: it opens only for large pools.

Study 1 simulates the Eyal-Sirer strategy, 2,000,000 blocks in each of 21 cells (q from 0.10 to 0.45, $\gamma \in \{0, 1/2, 1\}$). Without adjustment, the selfish pool's revenue per block interval never exceeds its honest revenue q (0 of 21 cells; at $q = 0.30$ and $\gamma = 0$ it earns 0.215 against 0.30). With adjustment, its share of the chain matches the closed form to within 0.001 and beats q in 14 cells, every one of them above the threshold.

Part II. The social layer

The results here vary in rigour, and each is labelled.

Where the proofs change kind

Remark (epistemic honesty). Bitcoin's security reduces to computational hardness assumptions, such as the preimage resistance of SHA-256, that are precisely defined, measurable in hash operations, and falsifiable by cryptanalysis. A vouching layer's security reduces to *social cost* assumptions that are hard to define precisely, not directly measurable, and dependent on human behaviour. This is a fundamental limitation, not a flaw to be hidden.

The design studied here is a vouching protocol with four parts: a new identity needs $k = 3$ existing members to vouch for it in person; a vouch carries consequences that ripple back to the vouchers if the new identity causes harm; the record of identities is hashed and anchored in Bitcoin; and verified members near an emergency can be asked to respond. The rule of three comes from the design's first statement: a token "only received if vouched in person by 3+ vouched peers". Gathering people in one place to issue one credential each has a history: pseudonym parties (Ford & Strauss, 2008) and proof-of-personhood gatherings (Borge et al., 2017). The related paper *Be in the Same Room at the Same Time* adds a synchronised in-person head count to the same idea and models its costs; here we analyse the vouching mathematics on its own.

The web-of-trust model

Definition 1 (verification graph). The verification graph is a directed graph $G = (V, E)$ in which V is the set of verified identities and $(u, v) \in E$ if u vouched for v. Each identity v has in-degree $d^-(v) \ge k$, with $k = 3$.

Assumption S (vouchers are present humans). Every vouch is given in person by a distinct verified human who is physically present, and each human can give at most m vouches.

Assumption S is what makes the counting below work, and it is where the physical layer earns its keep. Without it, k real vouchers seed one fake, and fakes then vouch for each other at no further social cost; this is exactly the attack that sinks purely online social-graph defences when attackers can make many cheap edges (Viswanath et al., 2010).

Assumption 1 (social cost exists). A voucher who takes part in a fraudulent verification incurs a cost $c_{\text{social}} > 0$: the risk of a lower trust score and of ejection (both quantifiable within the system), damage to their reputation among people they know (not quantifiable), and the time of an in-person meeting (quantifiable in hours). Unlike a hash cost, c_{social} differs from person to person and context to context; the honest statement is that it is a random variable with unknown distribution, and we assume $\mathbb{E}[c_{\text{social}}] > 0$.

Security Bound 1 (Sybil attack cost). Under Assumptions S and 1, creating n fake identities requires recruiting at least $\lceil nk/m \rceil$ distinct real vouchers, and if the attacker meets prospective vouchers at random,

$$\mathcal{C}(n) \ge \frac{nk}{m}\,\mathbb{E}[c_{\text{social}}].$$

Proof. n fakes need nk vouches. Under Assumption S none of them can come from a fake, and each real voucher gives at most m, so at least $\lceil nk/m \rceil$ distinct real people must be recruited. If the recruited people are drawn independently of their costs, the expected total cost is at least $\lceil nk/m \rceil\,\mathbb{E}[c_{\text{social}}]$. □

This is a lower bound, not the cost. Coordination, detection risk and the chance a recruit refuses or reports the attempt all add to it. With $k = 3$ and a lifetime limit of $m = 20$ vouches, 1,000 fakes need at least 150 real accomplices; with $m = 3$, 1,000 (Study 2).

The bound uses the *mean* cost, and that is its soft spot. An attacker who can choose whom to approach will start with the cheapest. Study 2 gives each of 100,000 reachable people a lognormal cost with mean 1 and lets the attacker recruit the cheapest. For 1,000 fakes, the cheapest 150 cost 18% of the bound when costs vary moderately (lognormal spread 0.5) and 2% when they vary widely (spread 1.0). The bound holds as stated for random encounters, and as $(nk/m)\times$ the mean cost of the cheapest recruits when vouchers can be chosen. Either way the cost still grows linearly with n, and the design's answer to selective recruitment is to make the cheapest vouchers expensive: the ripple and the separation rule (below) are both aimed at the person who vouches casually.

Open Question 1. Can c_{social} be measured? Three routes: the market price of fraudulent vouches if a black market appears; survey-based willingness to accept for vouching strangers; and revealed preference from attempted attacks. A survey design is in the repository's analysis plan.

Verification security

Assumption 2 (voucher independence). The k vouchers for a new identity are chosen independently, each colluding with probability p_c.

Security Bound 2 (fraud under independence). Under Assumption 2, $\Pr[\text{all } k \text{ collude}] = p_c^k$. With $k = 3$ and $p_c = 0.01$, the chance of a fraudulent verification is at most 10^{-6}.

Remark (the assumption is known to be optimistic). Colluders are correlated by definition: they coordinate. Study 2 gives each voucher a collusion propensity drawn from a shared beta distribution with mean 0.01 and intra-class correlation ρ (a beta-binomial model (Skellam, 1948)). A weak shared tendency, $\rho = 0.01$, lifts the three-way collusion rate from 1.0×10^{-6} to 5.9×10^{-6}; $\rho = 0.1$ lifts it to 2.1×10^{-4}, about 200 times the independent value. A more realistic model treats fraud as the existence of a colluding k-clique in the voucher graph, which depends on the graph's structure and is harder to analyse.

Heuristic 1 (separated vouchers). Require that no two of a new identity's vouchers are directly connected: $d(u_i, u_j) \geq 2$ for all $i \neq j$. A colluding group must then supply vouchers who look like strangers to one another, which raises the cost of coordination. The design specification adds a related path-diversity guard: reject a new identity if most of its sponsors share recent ancestry.

The heuristic's effect on a colluding group can be stated exactly for one model of the group.

Proposition 4 (admissible vouches in a colluding group). Let a colluding group of A people have internal ties forming a random graph in which each pair is connected independently with probability δ. Under the separation rule the expected number of admissible voucher sets of size k is $\binom{A}{k}(1-\delta)^{\binom{k}{2}}$; for $k = 3$, $\binom{A}{3}(1-\delta)^3$.

Proof. A set of k members is admissible exactly when none of its $\binom{k}{2}$ pairs is connected, which happens with probability $(1-\delta)^{\binom{k}{2}}$; linearity of expectation over the $\binom{A}{k}$ sets gives the result. □

A tight ring is the cheap ring to organise, and the rule shuts it out. In a group of 30 where half of all pairs know each other, only 12.4% of voucher triples are admissible; where 90% do, 0.1%; a group in which everyone knows everyone has none (Study 2, 400 simulated groups per cell, matching the formula). This is still a heuristic: a patient attacker can recruit people who do not know each other, and the rule says nothing about how expensive that is.

Ripple responsibility

Definition 2 (ripple responsibility). When identity v commits a violation of severity $s \in [0, 1]$, trust scores update as

$$\tau(u) \leftarrow \tau(u) - s\,\delta\,r^{d(u,v)},$$

where $d(u, v)$ is graph distance, $r \in (0, 1)$ is the decay rate and δ the base penalty.

Proposition 2 (bounded ripple, fixed decay). If every identity has at most Δ neighbours and $r\Delta < 1$, the total penalty from one violation is bounded:

$$\sum_{u \in V} \Delta\tau(u) \leq s\,\delta\,\frac{1}{1 - r\Delta}.$$

Proof. At distance i there are at most Δ^i identities, so the total is at most $s\delta \sum_{i \geq 0}(r\Delta)^i = s\delta/(1 - r\Delta)$, which converges if and only if $r\Delta < 1$. □

The count of identities at distance i can be tightened. Each identity at distance $i \geq 1$ is reached through one at distance $i - 1$, and every identity at distance $i - 1 \geq 1$ spends at least one of its Δ ties pointing back, so there are at most $\Delta(\Delta - 1)^{i-1}$ at distance i. The total is then at most $s\delta\,[1 + r\Delta/(1 - r(\Delta - 1))]$, which needs only $r(\Delta - 1) < 1$. On random graphs of up to 20,000 identities the largest observed totals sit under this sharper bound in every case, and on the ring ($\Delta = 2$) they meet it exactly (Study 3).

Remark (critical constraint). Version 2.0 noted that the convergence condition is restrictive: with $r = 1/3$ it requires $\Delta < 3$, so each person could vouch for at most two others, which is unrealistic for a growing network; setting $r = 1/(\Delta + \epsilon)$ restores convergence but weakens the incentive as the network grows. The sharper count relaxes this only to $\Delta < 4$. For a ripple that spreads in every direction through the graph, the remark stands.

The design does not spread it in every direction. Its specification sends the ripple to the identity's sponsors: one third to each of the three people who vouched it in, one ninth to each of *their* sponsors, and so on, to a default depth of four. Every member has exactly k sponsors (the vouches that admitted them), however many people they later vouch for. That changes the arithmetic completely.

Proposition 3 (the up-chain ripple). Let every non-founding member have exactly k sponsors, each of whom joined earlier, and let a violation by v cost each member u who sponsored v at chain distance d (directly, $d = 1$; a sponsor's sponsor, $d = 2$) a penalty $s\,\delta\,r^d$. Then:

1. the members at chain distance i carry at most $s\delta(kr)^i$ in total, so with $r = 1/k$ each generation of sponsors carries at most one unit, $s\delta$;
2. with the ripple capped at depth D and $r = 1/k$, the total penalty is at most $(D + 1)\,s\delta$;
3. uncapped, with $r = 1/k$ and N members, the total is less than $s\delta\,(\lfloor \log_k N \rfloor + 1 + k/(k-1))$;
4. with $r < 1/k$, the total is at most $s\delta/(1 - kr)$, whatever the size of the network.

Proof. Each member has exactly k sponsors, so at most k^i members are at chain distance i from v, giving (1) and (4) by summing a geometric series, and (2) by summing $D + 1$ terms of at most $s\delta$ each (distance 0 is v itself). For (3), at most N members are at any distance, so the layer at distance i carries at most $s\delta \min(1, Nk^{-i})$. With $L = \lfloor \log_k N \rfloor$ the first $L + 1$ layers contribute at most $L + 1$, and since $N < k^{L+1}$ the rest contribute less than $\sum_{j \geq 0} k^{-j} = k/(k-1)$. □

Items 1, 2 and 4 hold unchanged if a member reached along several sponsor paths is charged once per path, as the specification's layer weights can be read: there are at most k^i sponsor paths of length i, so each generation still carries at most $s\delta(kr)^i$. Item 3 needs each member counted once.

The design's own number, one third to each of three sponsors, is exactly the critical decay $r = 1/k$: the point at which each generation of sponsors carries the same total consequence, spread thinner, and no generation carries more. Three properties follow at once. The incentive does not weaken as the network grows: a direct sponsor bears one third of a violation in a network of a hundred or a hundred million. Nobody's liability is unbounded: the total from one violation is at most five units at the default depth of four. And how many people a member vouches for does not enter the bound at all. What it changes is how often a member is exposed, which is the point of vouching carefully, and the protocol's lifetime limit on vouches caps that.

Study 3 grows sponsor networks of 1,000 to 100,000 members, each newcomer choosing three earlier members as sponsors, either uniformly or from the 200 most recent joiners (a crude picture of local vouching). In every simulated violation each generation carried at most one unit; with the depth cap the largest total was exactly 5; uncapped, the mean total rose from 4.4 to 6.8 as the network grew a hundredfold, against a bound of 12.5 at 100,000 (Figure 1).

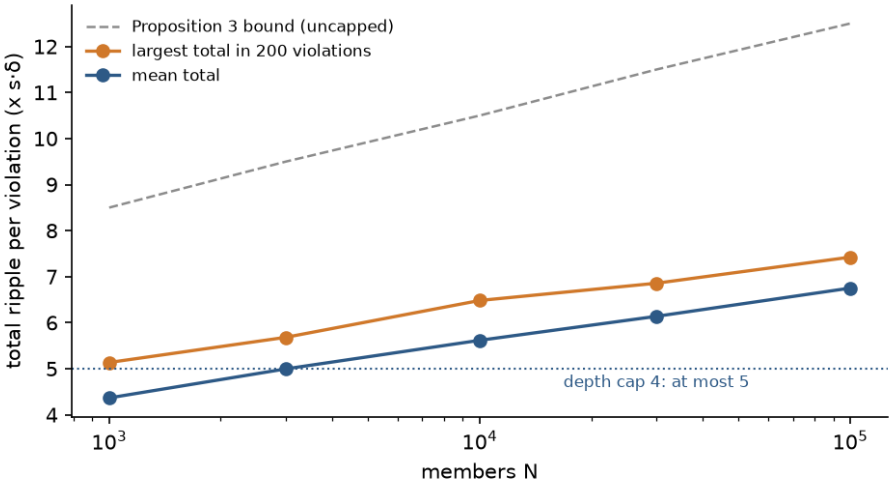


Fig. 1. Total ripple from one violation as the network grows, sponsor networks with three sponsors per member and decay one third (Study 3). Navy: mean of 200 simulated violations; orange: the largest; grey dashed: Proposition 3's uncapped bound; dotted: the design's depth cap of four, which holds the total to at most five.

Open Question 2, revised. Version 2.0 called the trade-off between decay rate and incentive strength "the hardest unsolved problem" in the trust model. For a ripple that runs up the sponsor chain with exactly k sponsors per member, Proposition 3 settles it: $r = 1/k$ keeps the direct incentive constant and the total bounded at every scale. What stays open is fairness. Should a sponsor who vouched in good faith carry the same third as one who vouched carelessly, and should the cap be four generations or fewer? Those are questions of justice, not convergence, and the mathematics cannot answer them.

Emergency response: an idealised model

Assumption 3 (homogeneous Poisson responders). Verified responders are spread as a homogeneous Poisson point process with intensity ρ responders per unit area.

Proposition 5 (coverage under the Poisson assumption). Under Assumption 3, the probability of at least one responder within distance d is

$$P_{\text{coverage}}(d) = 1 - e^{-\rho\pi d^2}.$$

Proof. The number of points in a disc of radius d is Poisson with mean $\rho\pi d^2$, and coverage fails only when it is zero. □

Simulated patterns match the formula to within 0.3 percentage points at every density tried (Study 4). At 25 responders per square kilometre, coverage within 200 metres is 95.7%.

Remark (critical limitation). Assumption 3 is unrealistic. Real populations cluster in cities, buildings and rooms; density changes through the day; walls and stairs slow real response; and responders are not always available or able. The "60-second response" target also assumes instant notification.

Version 2.0 conjectured that clustering lowers coverage below the Poisson value at the same mean density. That can be proved.

Theorem 6 (clustering never helps an independent place; was Conjecture 6.4). Let responders form any Cox process (a Poisson process whose intensity is itself random), with mean intensity ρ. This includes the Thomas cluster process (Chiu et al., 2013; Thomas, 1949). At a location chosen independently of the responders, the probability of at least one responder within distance d is at most the Poisson value $1 - e^{-\rho\pi d^2}$, with equality only if the random intensity over the disc is constant.

Proof. Let $\Lambda(B)$ be the random intensity measure of the disc B, so $\mathbb{E}[\Lambda(B)] = \rho\pi d^2$. Given Λ, the count in B is Poisson, so the chance of no responder is $\mathbb{E}[e^{-\Lambda(B)}]$. The function $x \mapsto e^{-x}$ is convex, so by Jensen's inequality $\mathbb{E}[e^{-\Lambda(B)}] \geq e^{-\mathbb{E}[\Lambda(B)]} = e^{-\rho\pi d^2}$, with equality only when $\Lambda(B)$ is almost surely constant. Coverage is one minus this. □

The same argument covers density that varies over a city in a known way: averaged over places, coverage is at most the Poisson value at the average density. Study 4 shows how large the gap gets. At 25 responders per square kilometre, Thomas clusters cut coverage at a random place from 95.7% to between 87.1% (small, loose clusters) and 38.7% (clusters of 25 within about 100 metres).

But incidents do not happen at random places. They happen to people, and people are where the clusters are. Study 4 builds a clustered population of 2,000 people per square kilometre (a suburban figure, illustrative), makes an independent 1-5% of them responders, and measures coverage at a randomly chosen person, counting only *other* responders. At 1% adoption, coverage at a person was 91.9% to 93.1% across cluster sizes against a Poisson value of 91.9%, while coverage at a random place fell to 80.3%. At 2% and 5% the figures at a person stayed within 1.5 percentage points of Poisson in every case. Clustering empties places that have no people in them; it barely changes coverage at the people. This model is ours and it is idealised (responders are a random share of residents, and the victim's own cluster is as likely to hold a responder as any other), but it suggests the Poisson figure is a fair guide for the question that matters.

Conjecture 1 (achievable coverage). In urban areas where adoption exceeds 1-5% of the population, several verified people will typically be within walking distance (200 metres) of most incidents. The binding constraint is not density but willingness: a verified person nearby means nothing if they do not respond.

Under the Poisson model at 2,000 people per square kilometre, 1% adoption puts an expected 2.5 adopters within 200 metres (two or more with probability 0.72), 2% puts 5.0 (0.96), and 5% puts 12.6. Willingness thins a Poisson process into another Poisson process, so the density needed for a target coverage is $-\ln(1-P)/(w\pi d^2)$ for a willing share w. If one responder in five answers, 95% coverage within 200 metres needs 119 adopters per square kilometre, 6% of residents at that density (Study 4, Figure 2). Coverage on real Australian densities, and the arrival times it implies, is the subject of the related paper *Civic Proximity Response*.

Remark (order of magnitude). The draft's working figure is that about one person in 4,600 needs an emergency response each day (one in 2,000 to 3,700 on other measures; its source is still being traced). A network of 100,000 members would then see about 22 to 50 emergencies a day. That is an engineering load, not a physical impossibility.

Cryptographic anchoring

Theorem 7 (inherited Bitcoin security). A record's root hash R_e, anchored in a Bitcoin transaction at height B and followed by z confirmations, can be reverted with probability

$$\Pr[\text{revert } R_e] \leq I_{4pq}\left(z, \tfrac{1}{2}\right),$$

where q is the attacker's share of Bitcoin's hashrate.

Proof. Reverting R_e requires replacing the block that contains it, which is a double-spend race against z confirmations; apply Theorem 2. □

Two of the claims register's entries follow directly. The cost of rewriting an anchored record is bounded below by the work of out-mining the honest chain from block B (the claim that attack cost is bounded by mining difficulty). And anchors inherit the

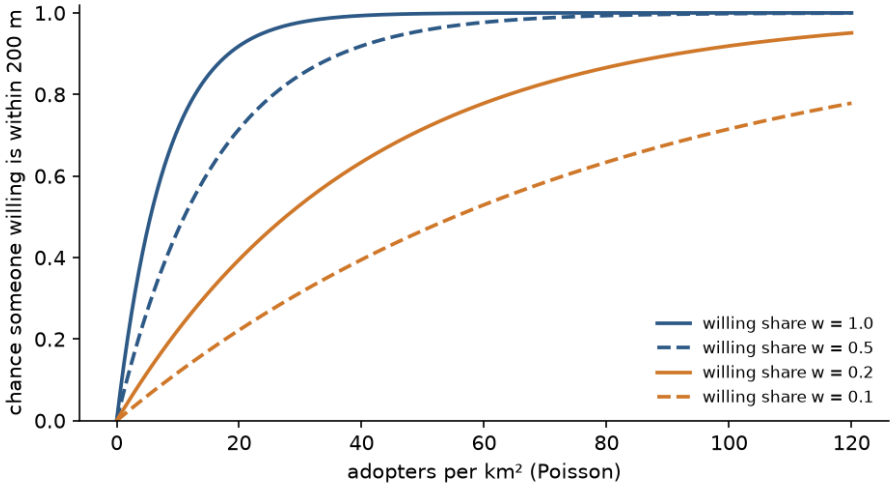


Fig. 2. Chance that someone willing to respond is within 200 metres, under the Poisson model, as adopter density rises, for willing shares from one in ten to all (Study 4). Willingness, not density, decides how much adoption is needed.

chain's total order of blocks (the claim that ordering is preserved), though not the accuracy of block timestamps, which miners set within a tolerance.

Table 3. Confirmations an anchored record needs before the chance of reverting it falls below one in a thousand, a million and a billion, by attacker share (Theorem 7; Study 1). Hours assume ten minutes a block.

Attacker share q	Risk below 10^{-3}	Below 10^{-6}	Below 10^{-9}	Hours to 10^{-6}
0.05	4	8	12	1.3
0.10	6	12	19	2.0
0.20	13	28	43	4.7
0.30	32	69	108	11.5
0.40	133	294	458	49.0
0.45	539	1,191	1,858	198.5

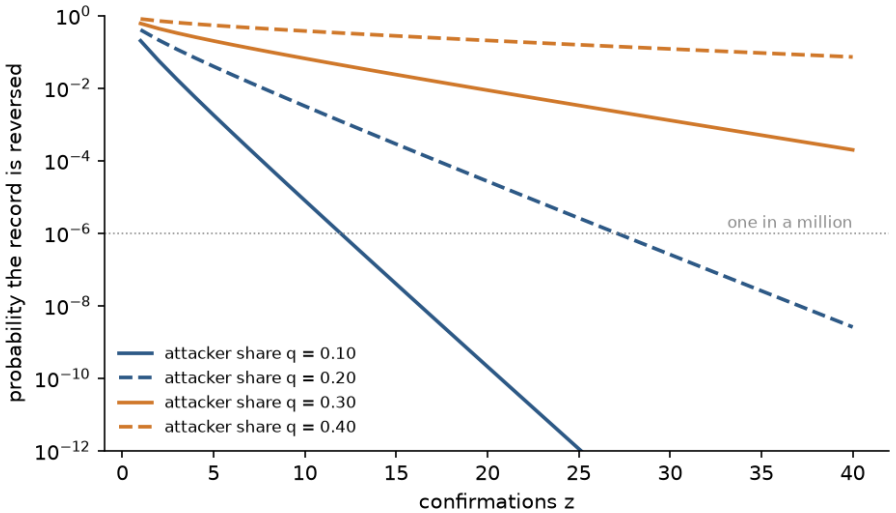


Fig. 3. Probability that an anchored record is reversed after z confirmations, exact formula, for attackers holding 10% to 40% of Bitcoin's hashrate (log scale; Study 1). The dotted line marks one in a million.

Theorem 7 is the only result on the social layer's side that is fully rigorous, because it borrows Bitcoin's proof whole. It protects the *record* of who vouched for whom; it says nothing about whether the vouches were honest. That is what Part II's bounds are for.

The signatures and storage the design uses rest on standard results, not on anything new here. BLS aggregate signatures are secure under the computational Diffie-Hellman assumption on their pairing groups in the random-oracle model (Boneh et al., 2004); Ed25519 signatures are high-speed Schnorr-type signatures (Bernstein et al., 2012) whose security has been proved in the random-oracle model (Brendel et al., 2021); and content addressing, as in IPFS (Benet, 2014), is sound so long as the hash function resists collisions. These are proven in the literature under standard assumptions, which is the most any deployed cryptography

Property	Bitcoin	Social vouching layer
Finality	Probabilistic, exact formula $I_{4pq}(z, 1/2)$	Social (three vouches), $\leq p_c^3$ under independence
What failure looks like	A SHA-256 break: obvious	A coordinated ring: possibly subtle
The record	Itself	Inherits Bitcoin's bound when anchored (Theorem 7)

Bitcoin's security fails loudly if it fails at all. A vouching layer's can fail quietly, one careless or bought vouch at a time. The design's answer is to make every vouch carry a consequence that scales (Proposition 3), to keep vouchers from being each other's friends (Heuristic 1, Proposition 4), and to put the record where nobody can rewrite it (Theorem 7).

Open problems and future work

1. **Measuring social cost** [high]. Can c_{social} be defined in measurable terms (time, money, reputation) and estimated? A willingness-to-accept survey is planned in the repository.
2. **Fairness of the ripple** [high]. Proposition 3 settles convergence and incentive for the up-chain ripple. Whether good-faith and careless sponsors should carry the same share, and how deep the ripple should go, remain open.
3. **Correlated collusion** [medium]. Security bounds for correlated vouchers, using dependent percolation or epidemic models on realistic social graphs (Mislove et al., 2007).
4. **Clustered coverage on real addresses** [medium]. Theorem 6 bounds coverage at places; the claim that coverage at people stays near the Poisson value rests on our model and should be tested on geocoded addresses.
5. **Empirical validation** [high]. Controlled experiments measuring the cost of Sybil attacks on in-person vouching.
6. **Adversarial analysis** [medium]. Attackers who exploit graph structure to minimise social cost, including recruiting mutual strangers to beat the separation rule.
7. **Privacy against verifiability; zero-knowledge at scale; cross-chain anchoring; cross-cultural trust** [open]. The claims register lists these as open or conjectured, and they stay so.

Limitations

Every number in this paper is computed from a model or a simulation of one. The Bitcoin studies assume the standard race model: an attacker who starts mining at the payment and miners who otherwise follow the protocol. The social-layer studies assume distributions for social cost and collusion that nobody has measured; they are swept to show which way the bounds move, not to estimate them. The sponsor networks of Study 3 are grown by simple rules, uniform and local, not taken from any real vouch graph. The populations of Study 4 are Thomas processes at a round illustrative density. The theorems hold under their assumptions whatever the data; the bounds and heuristics need data before anyone should rely on their numbers.

Conclusion

Bitcoin's security is proven. A vouching layer's security is argued, bounded and partly proven, and saying exactly which is which is itself a form of trustworthiness. This paper keeps that line where version 2.0 drew it and moves it in two places, both towards proof. The ripple of consequences, which the earlier version called its hardest unsolved problem, is bounded at every scale once it runs up the sponsor chain at one third per sponsor, which is how the design specifies it. And the conjecture that clustering hurts coverage is a theorem, with a turn the conjecture did not expect: clustering empties places, not the people who need help.

What a decentralised trust system can claim, then: Bitcoin-level security for the anchored record; a Sybil cost that grows linearly with the number of fakes, under stated assumptions; a ripple that keeps vouching careful without making anyone's liability unbounded; and coverage that depends on willingness more than density. What it cannot claim: Sybil resistance proven to the standard of Bitcoin's double-spend bound; a measured price for social trust; guaranteed emergency coverage. The open problems are listed. We invite anyone to strengthen these arguments or to find what we have missed.

Data and code availability

All four studies are in the paper's repository (`studies/`), with scripts and expected outputs; `reproduce.sh` reruns them and checks every number against `expected/`. The studies are verification of mathematics (models and simulations, seeded) and use no personal or empirical data. Version 2.0's double-spend table and its claims register are kept in the repository as records. The statistics appendix (`appendix/statistics-appendix.pdf`) sets out each check, why it fits, its assumptions and the sceptic's questions.

Related papers: *The Immutable Chain: How Cryptography Took Permission Out of Trust, from Caesar's Cipher to Bitcoin*; *Be in the Same Room at the Same Time: Physical Co-Presence as Sybil Resistance, and a Token System Built on It*; *Civic Proximity Response* (Applebee & Combe, 2026).

References

Benet, J. (2014). *IPFS: Content addressed, versioned, P2P file system*. arXiv preprint arXiv:1407.3561. https://arxiv.org/abs/1407.3561

Bernstein, D. J., Duif, N., Lange, T., Schwabe, P., & Yang, B.-Y. (2012). High-speed high-security signatures. *Journal of Cryptographic Engineering*, *2*(2), 77–89. https://doi.org/10.1007/s13389-012-0027-1

Boneh, D., Lynn, B., & Shacham, H. (2004). Short signatures from the Weil pairing. *Journal of Cryptology*, *17*(4), 297–319. https://doi.org/10.1007/s00145-004-0314-9

Borge, M., Kokoris-Kogias, E., Jovanovic, P., Gasser, L., Gailly, N., & Ford, B. (2017). Proof-of-personhood: Redemocratizing permissionless cryptocurrencies. *2017 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, 23–26. https://doi.org/10.1109/EuroSPW.2017.46

Brendel, J., Cremers, C., Jackson, D., & Zhao, M. (2021). The provable security of Ed25519: Theory and practice. *2021 IEEE Symposium on Security and Privacy (SP)*, 1659–1676. https://doi.org/10.1109/SP40001.2021.00042

Chiu, S. N., Stoyan, D., Kendall, W. S., & Mecke, J. (2013). *Stochastic geometry and its applications* (3rd ed.). Wiley. https://doi.org/10.1002/9781118658222

Douceur, J. R. (2002). The Sybil attack. *Peer-to-Peer Systems: First International Workshop, IPTPS 2002, Lecture Notes in Computer Science*, *2429*, 251–260. https://doi.org/10.1007/3-540-45748-8_24

Eyal, I., & Sirer, E. G. (2018). Majority is not enough: Bitcoin mining is vulnerable. *Communications of the ACM*, *61*(7), 95–102. https://doi.org/10.1145/3212998

Ford, B., & Strauss, J. (2008). An offline foundation for online accountable pseudonyms. *Proceedings of the 1st Workshop on Social Network Systems*, 31–36. https://doi.org/10.1145/1435497.1435503

Garay, J., Kiayias, A., & Leonardos, N. (2015). The Bitcoin backbone protocol: Analysis and applications. *Advances in Cryptology – EUROCRYPT 2015, Lecture Notes in Computer Science*, *9057*, 281–310. https://doi.org/10.1007/978-3-662-46803-6_10

Grunspan, C., & Pérez-Marco, R. (2018a). Double spend races. *International Journal of Theoretical and Applied Finance*, *21*(8), 1850053. https://doi.org/10.1142/S021902491850053X

Grunspan, C., & Pérez-Marco, R. (2018b). *On profitability of selfish mining*. arXiv preprint arXiv:1805.08281. https://arxiv.org/abs/1805.08281

Grunspan, C., & Pérez-Marco, R. (2020). The mathematics of Bitcoin. *EMS Newsletter*, *2020-3*(115), 31–37. https://doi.org/10.4171/NEWS/115/8

Grunspan, C., & Pérez-Marco, R. (2022). On profitability of Nakamoto double spend. *Probability in the Engineering and Informational Sciences*, *36*(3), 732–746. https://doi.org/10.1017/S026996482100005X

Lamport, L., Shostak, R., & Pease, M. (1982). The Byzantine generals problem. *ACM Transactions on Programming Languages and Systems*, *4*(3), 382–401. https://doi.org/10.1145/357172.357176

Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., & Bhattacharjee, B. (2007). Measurement and analysis of online social networks. *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, 29–42. https://doi.org/10.1145/1298306.1298311

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system*. https://bitcoin.org/bitcoin.pdf

Pass, R., Seeman, L., & Shelat, A. (2017). Analysis of the blockchain protocol in asynchronous networks. *Advances in Cryptology – EUROCRYPT 2017, Lecture Notes in Computer Science*, *10211*, 643–673. https://doi.org/10.1007/978-3-319-56614-6_22

Rosenfeld, M. (2014). *Analysis of hashrate-based double spending*. arXiv preprint arXiv:1402.2009. https://arxiv.org/abs/1402.2009

Sapirshtein, A., Sompolinsky, Y., & Zohar, A. (2017). Optimal selfish mining strategies in Bitcoin. *Financial Cryptography and Data Security 2016, Lecture Notes in Computer Science*, *9603*, 515–532. https://doi.org/10.1007/978-3-662-54970-4_30

Skellam, J. G. (1948). A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. *Journal of the Royal Statistical Society: Series B (Methodological)*, *10*(2), 257–261. https://doi.org/10.1111/j.2517-6161.1948.tb00014.x

Thomas, M. (1949). A generalization of Poisson's binomial limit for use in ecology. *Biometrika*, *36*(1/2), 18–25. https://doi.org/10.1093/biomet/36.1-2.18

Viswanath, B., Post, A., Gummadi, K. P., & Mislove, A. (2010). An analysis of social network-based Sybil defenses. *Proceedings of the ACM SIGCOMM 2010 Conference*, 363–374. https://doi.org/10.1145/1851182.1851226

Williams, D. (1991). *Probability with martingales*. Cambridge University Press. https://doi.org/10.1017/CBO9780511813658

Yu, H., Gibbons, P. B., Kaminsky, M., & Xiao, F. (2008). SybilLimit: A near-optimal social network defense against Sybil attacks. *2008 IEEE Symposium on Security and Privacy*, 3–17. https://doi.org/10.1109/SP.2008.13

Yu, H., Kaminsky, M., Gibbons, P. B., & Flaxman, A. (2006). SybilGuard: Defending against Sybil attacks via social networks. *ACM SIGCOMM Computer Communication Review*, *36*(4), 267–278. https://doi.org/10.1145/1151659.1159945

Author information

Affiliation: Omxus Press. All authors wrote the paper. The authors declare no competing interests. Correspondence: research@omxus.com.