Proof of Demand Is Not Proof of Work

On the Limits of Demand-Weighted Consensus under Free Pseudonyms

Ömer Demirel*

October 2026

Abstract

Proof-of-useful-work (PoUW) certifies computational hardness, not utility. We separate three properties of a work receipt: work soundness (W), job binding (B), and demand exogeneity (E), the last defined relative to an attacking coalition. Job binding is cryptographically achievable; demand exogeneity is not derivable from endogenous receipts. Payments between coalition-controlled requesters and workers are recoverable transfers, so no attack-cost lower bound may count their nominal value (Lemma 1). Under free pseudonyms and endogenous observation, receipts cannot certify E (Theorem 1); against a stated class of provenance estimators, evasion costs the equivalent of an implicit tax on simulated volume, below one percent at the parameters we consider. A mechanism that must stay safe and live on the zero-demand path needs an election path independent of demand-derived receipt weight (Proposition 4). The quantitative consequence is a sizing rule. A receipt boost that a coalition can realize by paying itself must be capped against the demand-independent floor (Corollary 4), and pricing the boost with a burn changes its cost but not this cap (Corollary 5): a burn-priced boost capped at κ times a participant's own stake lowers the floor share an attacker needs from τ to $\tau/(1+\kappa(1-\tau))$ once honest demand vanishes. In a stylized free-entry model, fork-independent salvage value of useful outputs leaves security neutral, lower, or higher depending on salvage asymmetry and demand. An irrecoverable tax on every settled payment makes the burn the security resource. We know of no production chain that weights leader election by endogenously attested demand, so the results constrain a design space; the deployed systems closest to it, Filecoin and Gridcoin, keep a demand-independent floor. Receipts suit payment, collateral, loss allocation, and a bounded, priced election boost. They cannot carry the leader-election floor.

1 Introduction

Proof-of-work (PoW) secures Nakamoto consensus by making the creation of consensus weight expensive and its verification cheap [1]. The expenditure is, by design, otherwise worthless. Proof-of-useful-work proposes to redirect it toward computation with external value—optimization, scientific computing, and, increasingly, machine-learning training and inference [12, 13, 19, 23]. The proposal is older than its solutions: a recent systematic review of fifty-three primary studies reports verification cost and difficulty adjustment as recurring obstacles for PoUW consensus [19].

This paper asks a narrower question than whether useful work is useful. Several designs propose to let paid service stand in for work: weight leader election by receipts for jobs that customers paid for, on the theory that real demand is costly to fake. The question is whether a receipt can show that its payment came from someone outside the attacking coalition. Reward programs that

*Snowfall Finance LLC. Revised preprint, v3.

weighted payouts by transaction volume have answered it in practice: self-dealt volume appears wherever volume is paid [9, 10], and airdrop programs have excluded farmed addresses by the thousands (§7). We study the same question for leader election. No production chain we know of yet weights election by endogenously attested demand, so the results constrain a design space before deployment. The deployed systems closest to it, Filecoin's committed-capacity floor and Gridcoin's stake-only block production, already keep a demand-independent floor (§7).

Two recent developments on the work side sharpen the contrast. On the constructive side, near-optimal work certification has been demonstrated for the matrix-multiplication primitive—$1 + o(1)$ multiplicative overhead over the native computation, asymptotically and under the construction's stated hardness assumptions [21]. Since matrix multiplication is a dominant computational kernel in modern machine-learning workloads, this substantially narrows the verification-asymmetry gap for that primitive, though not for complete inference graphs, stochastic generation, or general useful computation. Crucially, the guarantee is one of computational *hardness*, not computational *utility*: miners supply their own operand matrices, and usefulness is an economic hope rather than a cryptographic property—a limitation the construction itself acknowledges. On the empirical side, a preprint audit of Pearl's publicly released cuPOW mining paths reports that the most widely used released miner contains no inference path, and that a purpose-built miner submitting uniformly random matrices earned accepted shares across four hardware families [22]. The audit is not the discovery of a vulnerability. Random operands are accepted *by design*; what the audit measures is how completely that latitude was exercised in the audited paths.

The distinction is between certifying that a computation occurred and certifying that it was someone's demanded job. Pearl's reported failure is a job-binding failure ($\mathsf{B} = 0$): enforceable in principle, but unenforced in that design and repairable only by redesigning the verification relation. Demand exogeneity is different. No strengthening of execution proofs or job binding lets an endogenous receipt certify that its payment came from an economically independent demander. The architectural robustness result (Proposition 4) does not rely on indistinguishability. We require a permissionless mechanism to remain safe and live on the *zero-demand path*—an explicit adequacy criterion, premise P4, with an explicit escape for demand-dependent consensus—and require admissible executions to remain feasible after demand coordinates are set to their null values. Under those premises, the leader-election floor cannot be constituted by the *demand component* of service receipts, however genuine the demand behind them. External demand can add security while it lasts, but cannot supply the component that must survive its disappearance.

We decompose the security claims of a useful-work receipt into three distinct properties (pairwise non-implied; §5):

- **Work soundness (W):** producing a unit of consensus weight requires a lower-bounded *irrecoverable* computational cost.
- **Job binding (B):** the computation is bound to an authorized, committed, fresh job and delivered under its acceptance contract.
- **Demand exogeneity (E):** the requester's payment is economically independent of the worker and of any consensus coalition containing it; formally, it raises the coalition's net wealth (Definition 1).

W is a property of the weight-granting rule; B and E are properties of individual receipts, and E is relative to a coalition. B is cryptographically achievable. E, we show, is not derivable from endogenous receipts under free pseudonym creation and our observation model: a coalition controlling both requester and worker roles can produce receipts in which every cryptographic statement is true while every economic implication is false. The nominal payment is a *recoverable*

transfer; it contributes nothing to the cost of attack.

Contributions. (i) A receipt-level security model separating W, B, and E, with the mechanism-class and demand-environment definitions the results quantify over (§2). (ii) A zero-demand robustness result: safety and liveness on every zero-demand execution, together with null-demand projection closure, force an election path independent of *demand-derived* receipt weight; a freely realizable ungated boost is capped. Pricing a boost with a burn does not remove the cap: burn-priced and stake-coupled boosts are sized against the floor alone (Corollary 5). Under an additional scalar-cost model and optional minimality premise, full settlement–election separation is a design recommendation (§6). (iii) A demand-indistinguishability theorem in three variants (unlinkable provenance with coalition-realizable receipt fields; transparent ledger with mimicry cost, in statistical form and as an explicit reduction against efficient E-certifiers), establishing that endogenous receipts cannot prove E (§4). (iv) A recoverable-transfer lemma: nominal self-payments contribute no Sybil-resistant cost, so no security lower bound may count them; a net adversarial cost bound deducting *attacker-realizable, fork-independent salvage value*, extending [6]; and a static, frozen-supply free-entry characterization of when salvage is security-neutral, negative, or positive (§3; positioned as a complement to [25] (prior) and [26] (concurrent) in §9). (v) Two independent collapse modes—cheap-weight and usefulness collapse—and an attacker-side scalar-cost equivalence corollary. A constructive *payment-tax* corollary: an irrecoverable tax on every payment prices self-dealing and genuine demand identically, making the burn the security resource (§5). (vi) Deployed evidence: the reported Pearl cuPOW measurement as usefulness collapse (W=1 under the assumed cost floor, B=0, E=0) and incentive-program farming as a deployed *analogue* of demand-exogeneity failure (B=1, E=0) (§7).

The contribution claimed here is the framework and its assembly, not a new reduction technique: the W/B/E decomposition and six-direction witness table; recoverability-based weight accounting; the estimator-relative mimicry bound; the zero-demand premise discipline and escape map; the quantitative cap for freely realizable boosts; and the payment-tax construction. Lemma 1 is an internal-transfer identity, Theorem 1(c) is a standard certifier-as-distinguisher reduction, and Proposition 4 is an organizational robustness lemma under an explicit design norm, which is why this version labels it a proposition rather than a theorem. The sharpest quantitative content is the sizing rule that follows from it (Corollaries 4 and 5).

Main results, informally. *Lemma 1 (recoverable transfers).* A payment a coalition makes to itself changes its net wealth only by the transfer's nonrecoverable frictions, so nominal receipt volume buys no Sybil-resistant weight. *Theorem 1 (demand indistinguishability).* Under free pseudonyms, a coalition can reproduce the receipt distribution of economically independent requesters, so no endogenous receipt certifies E; evading a stated provenance-estimator class has a nonzero, lower-bounded cost (Proposition 3). *Proposition 4 (demand-independent floor).* A permissionless mechanism that must stay live when demand vanishes cannot let the demand component of receipts carry its election floor; a freely realizable boost is capped against the floor (Corollary 4), and so is a burn-priced one (Corollary 5). Proposition 4's proof uses only the decomposition and null-demand definitions of §2 and premises P4–P5; Lemma 1 motivates the accounting discipline but is not invoked by the projection proof. Sections 3–5 build the economic framework it organizes, and a reader wanting the architectural result first may read §2 and §6 directly.

The deployed record motivates a precise statement of the result: the reported Pearl audit found receipts with no job binding [22]; points and airdrop programs weighted rewards by synthesizable

volume at scale [9, 10]; and recent inference-consensus proposals leave proposer selection or its cost parameters underspecified (§7) [23]. We distinguish what follows from the stated premises from what remains empirical or design-dependent.

The practical consequence (§8) is that a receipt can prove work and delivery, but not economically independent demand. The demand component of receipts therefore cannot provide the necessary consensus floor, although a bounded receipt component may add security through an explicit burn or a separately proved non-scalar resource. A receipt offered as proof of demand carries no nonrecoverable cost of its own (Lemma 1) and, under free pseudonyms, cannot be upgraded into one (Theorem 1); it is not proof of work.

2 Model

Parties and identities. The system comprises *requesters*, who submit jobs and pay for them; *workers*, who execute jobs; *verifiers*, who check receipts; and a *settlement layer* that finalizes payments. Identities are public keys. We adopt the Sybil setting of Douceur [2]:

Assumption 1 (Free pseudonyms)**.** Any party can create arbitrarily many identities at negligible cost, and a coalition $\mathcal{A}$ may control identities in every role simultaneously. We treat a coalition as a single economic principal whose wealth is the sum over its identities (its members can make binding side payments). The assumption concerns identities, not principals: we do not assume that independent parties can be recruited into a coalition at zero cost.

Assumption 2 (Endogenous observation)**.** Consensus, and any mechanism deriving weight from receipts, observes only public artifacts: job commitments, payment transcripts, receipts, and chain state. No external identity, legal, or hardware trust root is available to the mechanism.

Jobs and receipts. A *job commitment* J is a signed structure binding: the principal (requester key), the program or model commitment, the input commitment, the execution environment, a chain-derived nonce or deadline, an *acceptance regime* t, and payment terms. A *receipt* ρ for J with output y, evidence π, delivery attestation d, acceptance artifact a, and settlement record s is *valid* iff

$$\mathsf{Valid}(\rho) \;=\; \mathsf{Auth}(J) \wedge \mathsf{Fresh}(J) \wedge \mathsf{Exec}(\pi, J, y) \wedge \mathsf{Deliver}(d, J, y) \wedge \mathsf{Accept}_t(a, J, y) \wedge \mathsf{Settle}(s, J, a). \tag{1}$$

Binding is *horizontal*: every clause of (1) is required in every acceptance regime; only the internal content of Accept_t varies. In all regimes, Accept_t asserts only that *the contractually designated adjudication process produced an acceptance artifact*; it is a process claim, not a semantic-correctness claim, and the distinction matters most in the non-contractible regime below.

Acceptance regimes and contractibility. The regime t is determined by the *contractibility* of the acceptance predicate, in the sense of incomplete-contract theory [11], applied to the job rather than to the producing technology:

- *Fully contractible:* the predicate is specified ex ante and evaluated deterministically (exact replay, execution proofs, a certificate checked by a verified acceptor).
- *Probabilistically contractible:* a statistical test, audit policy, and error bounds are specified ex ante (sampled re-execution of a deterministic trace with a soundness bound).

- *Non-contractible to the protocol:* the relevant quality cannot be completely encoded for automatic adjudication; enforcement is necessarily relational—reputation, repeated interaction, discretionary review, governance—and the acceptance artifact records a governance outcome.

Contractibility partitions acceptance predicates, not producer technologies: LLM-generated output may be fully contractible when acceptance is expressible ex ante (schema-valid JSON, tests passed, a proof accepted by a kernel, a numerical result within a certified error bound), while "persuasiveness" is observable to the parties yet non-contractible to the protocol. Nothing in our results depends on which regime a job occupies; the impossibility below is orthogonal to acceptance.

Costs. We write q for a coalition's absolute consensus weight and, against honest weight W_H, $q/(q+W_H)$ for its *share*; security statements concern the share required to defeat honest consensus, and we flag below where absolute and relative weight must not be conflated. For a coalition $\mathcal{A}$ acquiring weight q we distinguish: execution cost $C^{\mathcal{A}}_{\text{exec}}(q)$ (energy, hardware amortization, opportunity cost of compute); protocol fees or burns $F_{\text{burn}}(q)$ paid irrecoverably to the protocol or destroyed; capital cost $C_{\text{capital}}(q)$ of value subject to protocol-required lockup for the required duration; *mimicry cost* $C_{\text{mimic}}(q, \varepsilon)$, the *incremental* expense of making the self-dealt activity supporting weight q resemble a target provenance distribution to within statistical distance ε (routing, delay, counterparty diversity, capital float not already counted in C_{capital})—we write $C_{\text{mimic}}(q)$ when the tolerance is fixed by context; expected collectible penalties $\mathbb{E}[S(q)]$; and *attacker-realizable salvage value* $V^{\mathcal{A}}_{\text{salvage}}(q)$, the value $\mathcal{A}$ retains or can sell, net of liquidation cost and price impact, independently of whether its preferred history becomes canonical. Salvage is measured net of any market impact from disposing of the outputs, and net of the compute opportunity cost already carried in $C^{\mathcal{A}}_{\text{exec}}$; the two terms must not double-count the same foregone alternative use. A transfer is *recoverable* by $\mathcal{A}$ if it moves value between $\mathcal{A}$-controlled identities; it is *nonrecoverable* if it leaves the coalition (burns, fees to honest parties, collectible slashing).

Definition 1 (Demand exogeneity)**.** Fix a coalition $\mathcal{A}$. A settled payment of amount p to an $\mathcal{A}$-controlled worker is *exogenous with respect to* $\mathcal{A}$ if it raises $\mathcal{A}$'s aggregate wealth by p, net of any value $\mathcal{A}$ returns to the payer, directly or through intermediaries, within the accounting horizon. A receipt satisfies E *with respect to* $\mathcal{A}$ when its payment is exogenous with respect to $\mathcal{A}$, and satisfies E *for the mechanism* when it does so with respect to every admissible attacking coalition.

A payment from an $\mathcal{A}$-controlled identity is never exogenous with respect to $\mathcal{A}$, and a payment refunded to its payer through an intermediary is not either. E is a property of a (receipt, coalition) pair: the same receipt can satisfy it with respect to one coalition and fail it with respect to a larger one, which is why the security-relevant quantifier ranges over admissible attacking coalitions. W has a different type. It is a property of the weight-granting rule (what a unit of weight costs to produce), whereas B and E are properties of individual receipts; we say a receipt has W when it was produced under a mechanism with that property.

Consensus coupling. A *demand-weighted* mechanism assigns leader-election or reward weight as a function of valid receipts (their count, nominal value, or attested work). One lifecycle constraint is worth recording: if receipts influence election on the same chain that finalizes their settlement, the design is circular; either election consumes receipts finalized in a strictly earlier epoch, or settlement occurs through an independently finalized channel. We assume one of the two throughout.

Mechanisms, weight provenance, and demand environments. The results of §4 and §6 quantify over the following objects.

Definition 2 (Mechanism and transcript space)**.** A *mechanism* assigns, in each epoch, a leader-election weight function w over identities as a function of the public artifacts of Assumption 2 (receipts, chain state, burns, lockups). The *observable transcript space* is the space of those artifacts over an execution, with a receipt's fields as coordinates. Weight functions that read provenance features (funding graphs, timing, counterparty structure) are *provenance-sensitive*; those reading only receipt-internal content are *provenance-blind*.

Definition 3 (Demand environment and zero-demand path)**.** A *demand environment* is a process of external job arrivals and payments originating from parties economically independent of the participant set. The *admissible class* contains every such process, including the *zero-demand path*: zero external arrivals from genesis onward.

Definition 4 (Weight decomposition and precedence)**.** Partition the receipt fields read by the election rule into *demand coordinates* x_d (payment amount, economically meaningful requester-side content, and volume or count statistics of positively paid receipts) and *resource coordinates* x_r (attested nonrecoverable work, destroyed value, per-identity stake, burn, or lockup), so that $w = w(x_d, x_r)$. A zero-price job's program, input, or receipt count is not a demand coordinate merely because it appears in receipt syntax when weight is computed from its attested work.

For every feasible resource profile x_r, assume a canonical *null-demand completion* x_d^0: the same resource actions, with zero payment and no claim of external demand. Define the *demand-free weight*

$$w_r(x_r) := w(x_d^0, x_r).$$

The *demand-derived variation* is $\beta(x_d, x_r) := w(x_d, x_r) - w_r(x_r)$. The projection result uses only the defining equality above and therefore permits provenance-sensitive or nonmonotone rules. Results that call β a *boost*, including Corollary 4, explicitly restrict attention to profiles on which $\beta \geq 0$. We write π_r for the restriction replacing w by w_r.

Four precedence rules fix mixed cases. (i) A contribution factoring through destroyed value is burn-derived even when the burn is computed from a payment. (ii) A demand-enabled contribution capped by a committed resource is *resource-coupled*: for example, $w = W(1 + \min\{V, \kappa W\})$ has $w_r = W$ and profile-dependent boost cap $B(W) = \kappa W^2$. Raising that cap requires more work, but every boost below it may still be freely realizable at fixed W through self-payment. Resource coupling therefore changes the sizing function; it does not remove the sizing obligation. (iii) A fixed threshold that merely unlocks unbounded volume-derived weight is an eligibility gate, not a resource cap; the unlocked weight remains demand-derived. (iv) A nonnegative demand-derived boost is *freely realizable at resource profile x_r up to $B(x_r)$* if, holding the coalition's resource coordinates fixed, the coalition can realize every boost in $[0, B(x_r)]$ through self-authored demand at nonrecoverable cost bounded independently of the requested boost. It is *uniformly freely realizable* only when one execution-independent cap applies across the relevant resource profiles. Pure nominal self-payment with no volume-dependent burn has the first property by Lemma 1; a volume-dependent nonrecoverable fee does not. Resource-coupled forms generally have the first property with a profile-dependent cap, but not the second.

Definition 5 (Zero-demand necessity and election path)**.** A mechanism *depends on demand-derived weight on the zero-demand path* if safety or liveness fails on some admissible zero-demand execution under the restriction π_r (equivalently, when the demand-derived part $w - w_r$ is zeroed);

a resource is *necessary* for election if the mechanism depends on it in this sense. An *election path* is any restriction of the mechanism under which w loads only on a stated resource subset and consensus is safe and live; π_r is the election path loading only on resource coordinates.

Remark 1 (Scope of necessity). We define necessity relative to the zero-demand path rather than over all environments: the floor argument concerns robustness to the disappearance of demand, and dependence confined to demand-rich environments is not addressed by the necessity relation itself (a mechanism that loads on demand-derived weight only while demand is present, and on a receipt-independent resource when it is absent, satisfies the conclusion of §6). The quantitative cap of Corollary 4, however, does bind on demand-rich executions.

3 Recoverable transfers and the net cost of self-dealt work

Lemma 1 (Recoverable transfers)**.** *Under Assumption 1, any payment between a requester identity and a worker identity both controlled by coalition $\mathcal{A}$ changes $\mathcal{A}$'s net wealth only by the nonrecoverable frictions of the transfer. Consequently, nominal job payments, gross transaction volume, and aggregate receipt value between $\mathcal{A}$-controlled identities contribute zero net economic cost to $\mathcal{A}$, and no consensus-security lower bound may count them unless some portion is made nonrecoverable or the payer's independence is authenticated by an assumption external to the receipt system.*

Proof. Write $\mathcal{A}$'s wealth as the sum over its identities. An internal transfer of amount p from requester $r \in \mathcal{A}$ to worker $w \in \mathcal{A}$ adds $-p$ and $+p$ to the respective terms, leaving the sum unchanged up to fees ϕ leaving the coalition. The force of the lemma is an *optimization* statement, not a detection statement: a security lower bound must hold against the adversary's cost-minimizing strategy, and under Assumption 1 self-dealing—playing both roles and paying oneself—is always in the strategy set, at internal-transfer cost zero. Any claimed lower bound whose accounting includes p is therefore refuted by an available strategy that avoids p, overstating $\mathcal{A}$'s cost by exactly $p - \phi$; this holds *even if* common control of r and w were perfectly detectable, unless detection triggers a protocol response that makes some portion of the transfer nonrecoverable. Under Assumption 2, moreover, endogenous artifacts cannot establish common control with certainty. Graph-based *estimation* of common control remains possible and is treated in §4 and §8; estimation bounds risk, it does not restore the accounting. □

Lemma 2 (Net cost of the self-dealing strategy)**.** *Let a demand-weighted mechanism grant weight for valid receipts. Define the self-dealing strategy cost*

$$U_{\mathcal{A}}(q) \;=\; C^{\mathcal{A}}_{\text{exec}}(q) \;+\; F_{\text{burn}}(q) \;+\; C_{\text{capital}}(q) \;+\; C_{\text{mimic}}(q) \;+\; \mathbb{E}[S(q)] \;-\; V^{\mathcal{A}}_{\text{salvage}}(q), \qquad (2)$$

in which internal job payments do not appear (Lemma 1) and $V^{\mathcal{A}}_{\text{salvage}}(q)$ deducts the fork-independent value of the outputs the coalition performs for itself. The exhibited self-dealing strategy acquires weight q at expected net cost at most $U_{\mathcal{A}}(q)$; hence the minimal net cost $C^*_{\mathcal{A}}(q)$ of acquiring weight q satisfies $C^*_{\mathcal{A}}(q) \le U_{\mathcal{A}}(q)$. Consequently, no claimed attack-cost lower bound L for weight q may exceed $U_{\mathcal{A}}(q)$ (and no nonnegative bound may exceed $\max\{0, U_{\mathcal{A}}(q)\}$, the wrapper being slack exactly in the $U_{\mathcal{A}} < 0$ regime of Remark 6): an attack-strategy upper bound constrains how high the security floor can be; it is not itself a floor.

Proof sketch. The coalition plays both roles: it authorizes jobs to itself, executes them, delivers to itself, accepts, and settles internally. Each clause of (1) is satisfied by construction, so weight is

granted. The outlays are at most the nonrecoverable items listed; the internal payment cancels by Lemma 1. Because the coalition chose the jobs, it retains their outputs (trained weights, inference results, solved instances) whether or not any attack succeeds; their realizable value reduces net cost by $V^{\mathcal{A}}_{\text{salvage}}(q)$. For provenance-blind weight functions (§2) this naive strategy suffices with $C_{\text{mimic}} = 0$; provenance-sensitive weight functions are covered by pricing the required distribution matching into $C_{\text{mimic}}(q)$, the bridge made quantitative in §4. □

Proposition 1 (Salvage discount of useful election: unit cost)**.** *Fix a demand-weighted mechanism and compare it to an otherwise identical mechanism—same C_{exec}, F_{burn}, C_{capital}, C_{mimic}, $\mathbb{E}[S]$—whose work has no fork-independent value. By the accounting identity (2), attacker-realizable salvage lowers the exhibited self-dealing strategy's net cost of acquiring* absolute *weight q by exactly $V^{\mathcal{A}}_{\text{salvage}}(q)$ and therefore weakly lowers the minimal acquisition cost $C^*_{\mathcal{A}}(q)$. The reduction of the minimum is strict when $V^{\mathcal{A}}_{\text{salvage}}(q) > 0$ and the self-dealing strategy is cost-determining* in the no-salvage baseline*, or when every strategy minimizing cost in the baseline receives the same positive salvage discount; otherwise an already-cheaper strategy may leave the optimum unchanged.*

Proposition 1 prices absolute weight and holds unconditionally as accounting. It does not by itself price *security*, because salvage may also subsidize honest participation, raising the honest weight W_H that the attacker's share $q/(q + W_H)$ is measured against. The security-relevant statement requires an equilibrium model of honest supply.

Proposition 2 (Salvage and security: equilibrium form)**.** *Let each unit of weight cost c to produce, let honest participants realize salvage $v_H \in [0, c)$ and the coalition realize salvage $v_{\mathcal{A}} \in [0, c)$ per unit, let the protocol pay total reward flow R per epoch pro rata to weight holders, and let exogenous demand pay honest workers a flow $D \geq 0$ that the self-dealing coalition cannot capture. Under competitive free entry (entry until reward equals net unit cost) and holding the attacker's unit cost at $c - v_{\mathcal{A}}$ (no supply response to the attacker's own entry),*

$$W_H \;=\; \frac{R + D}{c - v_H}, \qquad C_{\text{match}} \;=\; W_H\,(c - v_{\mathcal{A}}) \;=\; (R + D)\,\frac{c - v_{\mathcal{A}}}{c - v_H}, \tag{3}$$

where C_{match} is the cost of acquiring weight equal to W_H (a 50% share). Honest supply is frozen at its pre-attack level; the addendum below restores honest exit and shows that C_{match} is an upper bound on the equilibrium matching cost. Equation (3) also credits salvage v_H to every demand-serving unit. That is exact when a non-rival output retains incremental value after delivery, but it overstates honest supply when the delivery contract makes delivery and retention alternative uses. Relative to the useless-work baseline $v_H = v_{\mathcal{A}} = 0$, $D = 0$ (where $C_{\text{match}} = R$), the comparison of C_{match} with R is governed by the single threshold

$$D \;\gtrless\; R\,\frac{v_{\mathcal{A}} - v_H}{c - v_{\mathcal{A}}}, \tag{4}$$

and, under free entry, the parameter space partitions into three mutually exclusive regimes by the sign of $D(c - v_{\mathcal{A}}) - R(v_{\mathcal{A}} - v_H)$:

(i) ***Neutral.*** *On the boundary of (4)—in particular at symmetric salvage ($v_{\mathcal{A}} = v_H$) with $D = 0$—$C_{\text{match}} = R$: the attacker's per-unit discount is exactly offset by expanded honest participation.*

(ii) ***Negative.*** *If $v_{\mathcal{A}} > v_H$ and $D < R(v_{\mathcal{A}} - v_H)/(c - v_{\mathcal{A}})$, then $C_{\text{match}} < R$: attacker-asymmetric salvage lowers the cost of matching honest weight whenever demand inflow is below the threshold (in particular at $D = 0$ for any $v_{\mathcal{A}} > v_H$).*

(iii) ***Positive: honest-favoring salvage, or above-threshold demand.*** *If* $D > R(v_{\mathcal{A}} - v_H)/(c - v_{\mathcal{A}})$ *then* $C_{\text{match}} > R$*, and the two ways this arises differ in persistence. When* $v_{\mathcal{A}} < v_H$ *the threshold is strictly negative, so* $C_{\text{match}} > R$ *for all* $D \geq 0$*,* including $D = 0$*: honest-favoring salvage raises the matching cost permanently, with no demand required. When* $v_{\mathcal{A}} \geq v_H$ *the threshold is nonnegative, so positivity requires strictly positive, above-threshold demand and is genuinely demand-contingent, holding only while* D *exceeds the threshold. (The permanent case is captured by the work/burn classification of Definition 4; the demand-contingent case is the one Proposition 4 removes from the floor, as Remark 4 discusses.)*

For risk-neutral participants, regime membership (though not the numerical values) is invariant across all weight-symmetric *reward-division conventions—those paying every unit of weight the same expected reward per epoch—as shown in the proof. Figure 1 displays the partition for fixed* (R, c, v_H)*. Separately from this partition, if honest supply is* inelastic*—a binding constraint* $W_H < R/(c - v_H)$ *with* $D = 0$ *that replaces the free-entry display (3) rather than adding a fourth cell—then for* $v_{\mathcal{A}} \geq v_H$*,* $C_{\text{match}} = W_H(c - v_{\mathcal{A}}) < R$ *already at* $v_{\mathcal{A}} = v_H$*: this is driven by the supply friction, not by usefulness (it holds even at* $v_{\mathcal{A}} = v_H = 0$*), and salvage asymmetry only widens the gap.*

Proof. Free entry equates marginal reward $(R + D)/W_H$ with net unit cost $c - v_H$, giving (3); the attacker matching W_H pays $c - v_{\mathcal{A}}$ per unit by Lemma 2 with salvage realized fork-independently. This acquisition-cost accounting does not credit the attacker with its pro-rata protocol reward: that reward is fork-dependent and is devalued on a consensus-breaking path, as Remark 5 models explicitly. Then

$$C_{\text{match}} - R \;=\; \frac{(R + D)(c - v_{\mathcal{A}}) - R(c - v_H)}{c - v_H},$$

whose sign is the sign of $D(c - v_{\mathcal{A}}) - R(v_{\mathcal{A}} - v_H)$; the threshold (4) and regimes (i)–(iii) follow, and they are exhaustive and mutually exclusive by trichotomy of that sign. For the separate inelastic case, the binding constraint replaces the free-entry display and, for $v_{\mathcal{A}} \geq v_H$, $C_{\text{match}} = W_H(c - v_{\mathcal{A}}) \leq W_H(c - v_H) < R$ directly. The equalities are exact under the stated idealization (pro-rata reward division; no honest supply response to attacker entry). We do not claim the numerical values are convention-free. Regime membership, however, is invariant over a delimited family: call a reward-division convention *weight-symmetric* if every unit of weight earns the same expected reward per epoch (pro-rata payment, a proportional block-winner lottery, and any mixture of such rules; see [45] for an axiomatic treatment of such rules). For risk-neutral participants, free entry under any weight-symmetric convention equates the same expected per-unit reward $(R + D)/W_H$ with the same net unit cost $c - v_H$, reproducing (3) and hence the sign of $C_{\text{match}} - R$ exactly; the regime partition is therefore identical across the family. Outside it—conventions paying different expected rewards to different units, such as superlinear leader bonuses—we make no invariance claim.

Honest exit (the equilibrium object). Display (3) freezes W_H at its pre-attack level and hence overstates the matching cost; restoring free *exit* gives the equilibrium object. With the attacker holding share σ, the reward R is split pro rata (the honest side keeps $R(1 - \sigma)$) while D accrues to honest weight alone, so free entry pins $W_H(\sigma) = [R(1-\sigma)+D]/(c-v_H)$. At a 50% share ($\sigma = \frac{1}{2}$) this gives $W'_H = (R/2+D)/(c-v_H)$ and $C'_{\text{match}} = W'_H(c-v_{\mathcal{A}}) = (R/2+D)(c-v_{\mathcal{A}})/(c-v_H) \leq C_{\text{match}}$. The absolute gap, $\frac{R}{2}(c-v_{\mathcal{A}})/(c-v_H)$, is independent of D but depends on the salvage parameters; the proportional reduction, $(C_{\text{match}} - C'_{\text{match}})/C_{\text{match}} = R/[2(R + D)]$, is independent of the salvage parameters but depends on D. (The naive reading that free entry pins *total* weight at

$(R + D)/(c - v_H)$ fails once the attacker—who cannot capture D—is present: total weight is $(R + 2D)/(c - v_H)$, collapsing to Budish's [3] one-for-one crowd-out only at $D = 0$.) Against the consistently re-equilibrated useless-work baseline $R/2$, $C'_{\text{match}} - \frac{R}{2} = [D(c - v_{\mathcal{A}}) - \frac{R}{2}(v_{\mathcal{A}} - v_H)]/(c - v_H)$ has the same trichotomy with $R \mapsto R/2$. The negative region may shrink when $v_{\mathcal{A}} > v_H$ because its boundary halves; the pointwise fact needed here is only $C'_{\text{match}} \leq C_{\text{match}}$, with strict inequality for $R > 0$.

Delivery channel. Equation (3) credits salvage v_H to every honest unit. Where the delivery contract is exclusive—for example, confidentiality or transfer of sole possession—delivery and retained salvage are alternative uses, and removing the double count weakly lowers honest supply and the attacker's matching cost. Where an information good is non-rival, delivery and retained value may coexist; v_H should then include only the incremental residual value that remains after satisfying the customer. The exact correction depends on which producer type is marginal and on how demand is allocated, so no universal amount such as $v_H W_D/(c - v_H)$ is asserted without a two-type entry model. Any correction vanishes at $D = 0$, leaving the zero-demand classification intact. □

Remark 2 (What is held fixed). Proposition 2 holds the protocol reward flow R fixed and lets free entry determine honest weight. Its neutral symmetric-salvage case is not invariant to another normalization. If instead honest weight is fixed at $W_H = R_0/c$, symmetric salvage is $v \in (0, c)$, and the designer reduces the reward to preserve that weight, then $R_1 = W_H(c - v) = R_0(c - v)/c$ and the matching cost is also $R_0(c - v)/c < R_0$. At equal physical honest weight, symmetric salvage therefore reduces dollar attack cost by the factor $(c - v)/c$. This does not contradict the fixed-R result: salvage buys more physical weight for a given emission, but banking that gain as a lower emission reduces the dollar security budget one-for-one.

Remark 3 (Reduced form: demand follows weight). Proposition 2 pays demand D pro rata to honest consensus weight. That is the right reduced form when each unit of weight is one unit of service capacity and customers spread jobs across capacity. In a service market, customers choose providers by price and quality, and in a demand-weighted mechanism causality runs the other way: serving demand produces weight. Two consequences follow. First, D accrues to the providers that win the service market rather than to marginal consensus weight, so the free-entry conditions for consensus weight and for service capacity can differ. Second, the hypothesis that the coalition cannot capture D excludes the case in which the coalition is itself a large provider; Remark 6 treats that case separately, at remark strength. Turning either point into a theorem requires a two-type entry model, with service providers and consensus participants and demand allocated by price, which we leave open. We therefore read Figure 1 as a map of the reduced form and rely on it only for the zero-demand classification, which the objection does not touch because $D = 0$ there.

Remark 4 (Which regime is operative: a positioned conjecture). Proposition 2 supplies sufficient conditions, not an empirical verdict, and adjacent equilibrium analyses reach the neutral case by different routes. Pass [26], concurrent with this work, shows in a competitive-equilibrium model of PoUW with *exogenous* inference demand and price adjustment that the economic cost of majority attack remains tied to the block reward across all regimes—neutrality as an equilibrium outcome, not a knife-edge. Bar-On, Komargodski, and Weinstein [25] (prior work) analyze heterogeneous-cost mining with external utilities and reach only a *conditional* decentralization verdict: their own analysis flags that concentrated access to external utilities can let a single actor control the network cheaply and that low coupon entropy reduces security. Our analysis is the attack-cost-side complement, and it opens three channels these models do not price: demand may be *endogenous* (self-dealt, Lemma 1), salvage may be attacker-asymmetric, and honest supply may be constrained.

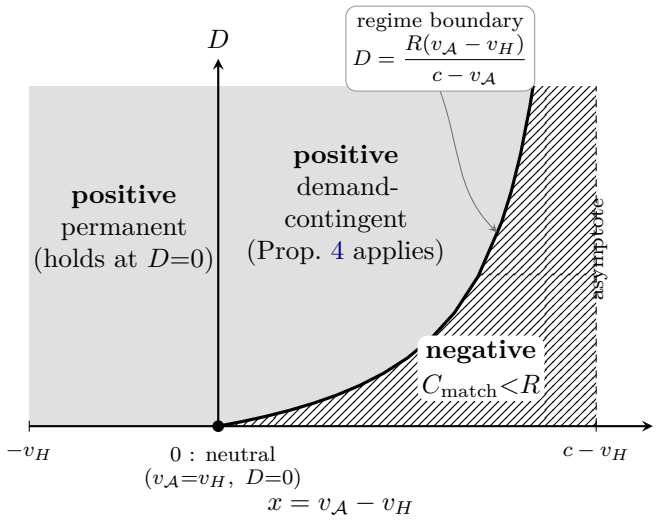


Figure 1: Salvage/security regime map at fixed (R, c, v_H) (Proposition 2). Horizontal axis: salvage asymmetry $x = v_{\mathcal{A}} - v_H \in (-v_H,\ c - v_H)$; vertical axis: exogenous demand flow $D \geq 0$. The neutral boundary $C_{\text{match}} = R$ is the *curve* $D = R(v_{\mathcal{A}} - v_H)/(c - v_{\mathcal{A}})$ (equivalently $D = Rx/(c - v_H - x)$), through the origin with a vertical asymptote at $x = c - v_H$; it bends because the denominator $c - v_{\mathcal{A}}$ moves with $v_{\mathcal{A}}$. For $v_{\mathcal{A}} < v_H$ (all of $x < 0$), $C_{\text{match}} > R$ for every $D \geq 0$, including $D = 0$: honest-favoring salvage is *permanently* security-positive. For $v_{\mathcal{A}} > v_H$ the plane splits—below the curve $C_{\text{match}} < R$ (negative), above it $C_{\text{match}} > R$ but only while demand exceeds the threshold (*demand-contingent* security, the component removed from the floor by Proposition 4).

Within that scope the conditional claim is exactly Proposition 2(ii); the concentrated-access threat Bar-On identify is, in our vocabulary, the paid-attacker centralization of Remark 6, to which our answer is the demand-independent floor rather than a liquidity assumption. Whether deployed systems occupy the asymmetric, inelastic regimes is an *empirical conjecture*: an entity capable of mounting a consensus attack plausibly has better resale channels and internal consumption for useful outputs than the marginal honest participant, and specialized useful-work supply faces attestation, hardware, and integration frictions. Two caveats bound the conjecture. At attack scale, per-unit salvage is plausibly *concave* (liquidation cost and price impact, already netted in $V^{\mathcal{A}}_{\text{salvage}}$'s definition), which weakens the asymmetry channel exactly where it matters most; and under demand-funded participation, case (iii) applies while the demand persists. Proposition 4, rather than this conjecture, supplies the response to case (iii): demand-contingent security cannot be the floor.

Remark 5 (Fork-independent salvage). Salvage enters (2) at face value only because it is *fork-independent*, and this is the property that removes it from Budish's [3] deterrence bound. Consider an attack path on which token-denominated, fork-*dependent* value is scaled by a devaluation factor $\delta \in [0, 1)$ (the post-attack collapse), while fork-independent salvage is unaffected. Split the attacker's per-unit proceeds into a fork-dependent part v_{dep} (self-paid tokens, protocol rewards, onchain resale) and the fork-independent part $v_{\mathcal{A}}$, giving attack-path unit cost $c - \delta v_{\text{dep}} - v_{\mathcal{A}}$. Two limits bound the picture. If all attacker salvage were fork-dependent ($v_{\mathcal{A}} = 0$), the matching cost $W_H(c - \delta v_{\text{dep}})$ rises to $W_H c$ as $\delta \to 0$—the useless-work acquisition cost at fixed honest weight—so the collapse restores deterrence. With fork-independent salvage (the PoUW case), the $v_{\mathcal{A}}$ term survives every δ, so the collapse deters nothing on the salvage margin and the negative regime of

Proposition 2(ii) persists under arbitrarily severe post-attack devaluation. Fork-independence is thus not a modeling convenience but the exact hypothesis under which useful-output value may be deducted from an attack-cost bound—and the reason the output-uselessness of conventional PoW, which forces $v_{\text{dep}} = v_{\mathcal{A}} = 0$, pins it to the δ-robust neutral case.

Remark 6 (Negative net cost: the paid attacker). Nothing in (2) forces $U_{\mathcal{A}}(q) \geq 0$. When genuine external demand is large and concentrated, a *dominant honest provider* may find $U_{\mathcal{A}}(q) < 0$: it is paid—by real, independent customers—to perform work that simultaneously accumulates consensus weight. Formally, this reads external customer revenue as realized, fork-independent value under $V^{\mathcal{A}}_{\text{salvage}}$'s definition—an *asserted* accounting extension, natural when service payments settle through the independently finalized channel of the lifecycle constraint (§2), but not derived within Proposition 2, whose coalition by hypothesis cannot capture D; we state the observation at remark strength. No fake demand, no self-dealing, and no violation of E is involved; the mechanism itself converts commercial success into consensus power. The immediate result is concentration of election power and the associated liveness, censorship, and governance risk. It is not by itself a safety attack: the provider's customer franchise, protocol rewards, and token holdings may be fork-dependent and lost if it attacks, so Budish-style deterrence can remain strong even when acquisition cost is negative. Safety is threatened only when the benefit of deviation exceeds those lost business and asset values. The demand-independent floor limits the control conferred by commercial concentration without assuming that the provider will attack.

Remark 7 (Why useless output matters). The output-uselessness of conventional PoW is not incidental: it keeps the fork-independent salvage value of mining near zero, secondary channels such as merged mining and heat reuse notwithstanding—pinning the system to the baseline $v_H = v_{\mathcal{A}} = 0$ of Proposition 2, where the neutral case holds trivially and no asymmetry can open. Dotan and Tochner [6] showed computationally that private information, easy instances, and external work markets undermine useful-work security, and that secure user-uploaded work collapses toward preimage-resistant tasks; Lemma 2 and Propositions 1–2 are the economic extension: even *honestly hard, honestly executed* useful work subsidizes the adversary that performs it for itself, and the subsidy is security-degrading in the asymmetric and inelastic regimes of Proposition 2(ii).

4 Demand indistinguishability

Job binding establishes *what computation served*. Demand exogeneity establishes *whether an independent party paid for it*. The next theorem shows that the second cannot be inferred from the first, or from any endogenous artifact.

Theorem 1 (Demand indistinguishability)**.** *Consider any protocol satisfying Assumptions 1 and 2 in which payments are fungible and coalitions may control requester and worker roles simultaneously.*

(a) ***Unlinkable provenance with realizable fields.*** *Suppose payment and requester artifacts reveal no linkable provenance, and,* for each target *economically independent requester–worker pair, the coalition can efficiently* realize *that pair's distribution over all remaining publicly observable receipt fields on the transcript space of §2—job metadata, timing, program and model choices, outputs, and acceptance artifacts—by executing the committed jobs and invoking the same endogenous verification and adjudication procedures available to honest participants; any fees owed to independent verifiers or adjudicators are nonrecoverable and accounted in C_{mimic} or an explicit fee term. Then for every transcript distribution produced by an economically independent requester–worker pair, a coalition $\mathcal{A}$ controlling both roles*

can produce an identically distributed transcript while recovering all internal transfers, at net cost at most $U_{\mathcal{A}}(q)$. Realizability, not mere knowledge of the artifact distribution, is what the premise requires: sampling the bytes of a signed acceptance artifact does not forge its signature. The variant is premise-conditional by design; its scope is discussed in Remark 8.

(b) ***Transparent ledger (statistical form).*** *Suppose provenance is public, and let $\mathcal{D}_H$ be the honest distribution over full transcripts on the observable space of §2 (all receipt and provenance features jointly). If $\mathcal{A}$ can produce a transcript distribution $\mathcal{D}_{\mathcal{A}}$ with statistical distance $\Delta(\mathcal{D}_{\mathcal{A}}, \mathcal{D}_H) \leq \varepsilon$ at incremental cost $C_{\text{mimic}}(\varepsilon)$, then every distinguisher—regardless of computational power—distinguishes self-dealt from exogenous demand with advantage at most ε, and $\mathcal{A}$'s net cost is at most $U_{\mathcal{A}}(q)$.*

(c) ***Transparent ledger (computational form).*** *Suppose provenance is public and transcripts are indexed by a security parameter λ. If $\mathcal{A}$ can produce a transcript ensemble computationally indistinguishable from the honest one—every probabilistic polynomial-time (PPT) distinguisher has advantage at most $\varepsilon(\lambda)$—then no PPT* E*-certifier is simultaneously complete and sound: any PPT procedure accepting honest-demand transcripts with probability at least $1 - \delta_c(\lambda)$ and accepting coalition transcripts with probability at most $\delta_s(\lambda)$ satisfies $\delta_c(\lambda) + \delta_s(\lambda) \geq 1 - \varepsilon(\lambda)$ for every λ. The statistical and computational notions are not mixed: (b) bounds all distinguishers under statistical distance; (c) bounds efficient ones under computational indistinguishability.*

In all three cases, receipt contents alone cannot establish E*: every clause of (1) can hold with* $\mathsf{E} = 0$.

Proof sketch. (a) The simulator is direct: $\mathcal{A}$ generates fresh requester keys, funds them from coalition capital, realizes job metadata, timing, program choices, and adjudication outcomes by running the committed jobs through the same endogenous procedures honest participants use, signs well-formed job commitments with chain-derived nonces, executes, delivers, produces acceptance artifacts under the declared regime, and settles. Each predicate in (1) is satisfied by an authentic cryptographic statement—the authorization *is* a valid signature, the freshness *is* a fresh nonce. Fix a single target independent pair. By per-target realizability the coalition reproduces that pair's field distribution, and unlinkability removes provenance as a distinguishing feature, so the simulated single-pair transcript is identically distributed to the honest one on all observable fields; internal payments cancel by Lemma 1. The claim is per pair and does not assert a match of the *joint* distribution across pairs: where honest receipts carry cross-pair correlation the coalition cannot reproduce (correlated arrival times, shared exogenous inputs), that structure is not realizable by hypothesis, and the relevant statement is variant (b), with the residual distance priced at $C_{\text{mimic}}(\varepsilon)$—the routing Remark 8 makes informally. (b) With public provenance, honest and simulated transcripts are random variables over the same observable space; any distinguisher's advantage is bounded by $\Delta(\mathcal{D}_{\mathcal{A}}, \mathcal{D}_H) \leq \varepsilon$ by the standard property of statistical distance. The cost of achieving ε-closeness—external capital float, routing depth, counterparty diversity, and delay—is $C_{\text{mimic}}(\varepsilon)$ by definition. (c) The reduction is direct: an E-certifier is itself a distinguisher. Given a PPT certifier C with completeness $\Pr[C(\Gamma_H) = 1] \geq 1 - \delta_c$ and soundness $\Pr[C(\Gamma_{\mathcal{A}}) = 1] \leq \delta_s$, the distinguisher that outputs "honest" exactly when C accepts has advantage $\Pr[C(\Gamma_H) = 1] - \Pr[C(\Gamma_{\mathcal{A}}) = 1] \geq (1 - \delta_c) - \delta_s$ (when $(1 - \delta_c) - \delta_s \leq 0$ the claimed inequality holds trivially, so assume it positive). Since C is PPT and the ensembles are $\varepsilon(\lambda)$-indistinguishable to PPT distinguishers, $(1 - \delta_c) - \delta_s \leq \varepsilon(\lambda)$, i.e. $\delta_c + \delta_s \geq 1 - \varepsilon(\lambda)$: any efficient certifier fails completeness, fails soundness, or splits the failure between them. □

Remark 8 (Scope of the realizability premise). Theorem 1(a) depends on its realizability premise. Realizability *holds* where the observable receipt distribution is a function of quantities the coalition controls or can execute: self-chosen job metadata, program and model commitments, timing patterns, outputs of jobs the coalition actually runs, and acceptance artifacts produced by endogenous adjudication open to all participants. It *fails*, or becomes expensive, where honest receipts embed structure correlated with private off-chain state the coalition cannot cheaply reproduce: inputs derived from proprietary datasets whose statistics differ from synthetic ones; demand arrival processes correlated with external observables (market hours, news flow, seasonal load) that a simulator must track; long-lived requester histories spanning venues the coalition does not control; or acceptance regimes requiring interaction with independent parties whose fees are nonrecoverable. In those cases exact simulation degrades to approximate simulation, and Theorem 1(b) applies with the residual distance ε priced at $C_{\mathrm{mimic}}(\varepsilon)$. Thus either the receipt distribution is coalition-realizable and demand is unprovable, or the unrealizable components are the imported exogenous structure of Remark 10.

Remark 9 (What survives: estimation, not proof). On transparent ledgers, endogenous evidence cannot certify E soundly and completely; transparent provenance can *falsify* independence in some cases (requester and worker sharing a key), and *estimate* it in others, but absence of detected linkage does not establish independence. The estimation is real: the empirical wash-trading literature detects self-dealt volume from exactly the provenance features above [9, 10]. Theorem 1(b) formalizes the arms race: an improving estimator raises $C_{\mathrm{mimic}}(\varepsilon)$ but never converts an estimate into a cryptographic guarantee. We deliberately name the term *mimicry* rather than *laundering*: legitimate privacy-preserving users produce similar patterns, and the quantity is a distribution-matching cost, not a moral category.

The estimation also has a *lower bound*, which keeps C_{mimic} from functioning as an unfalsifiable free parameter: against a stated class of depth-and-dwell provenance estimators, evading detection forces the coalition to immobilize capital as float, at a cost that scales with simulated volume and the estimator's parameters rather than with the work performed. The next proposition makes this precise, converting C_{mimic} from a posited constant into a bound relative to the estimator class; the point is the form, not any single number, and raising the estimator's parameters is exactly the design lever of §8.

Proposition 3 (Mimicry lower bound against depth/dwell estimators)**.** *Fix $k \in \mathbb{N}$ with $k \geq 1$ and a float threshold $\tau_{\mathrm{fl}} > 0$, and let the deployed provenance estimator belong to the class $\mathcal{E}(k, \tau_{\mathrm{fl}})$: estimators that flag any requester whose on-ledger funding path, traced from a source the estimator recognizes as $\mathcal{A}$-linked, either* (i) *shows fewer than k hops or* (ii) *contains a hop preceded by dwell time below τ_{fl}. Suppose a coalition $\mathcal{A}$ simulates n economically independent requesters, each transacting volume P funded from capital whose source the estimator already recognizes as $\mathcal{A}$-linked, and assume every funding path available for this capital begins at such a recognized source. If every simulated requester evades every estimator in $\mathcal{E}(k, \tau_{\mathrm{fl}})$, then each requester's P is immobilized for time $\geq k\tau_{\mathrm{fl}}$; hence, at capital-cost rate r (value per unit value per unit time),*

$$C_{\mathrm{mimic}} \;\geq\; nP\,k\,\tau_{\mathrm{fl}}\,r.$$

This bound needs no concurrency premise, because committed capital-time adds across paths even when the same capital is reused in sequence. If, in addition, all n requesters must be concurrently active within a window of length $\Omega < k\tau_{\mathrm{fl}}$ (for example, because their weight must be realized within one attack window) and value at float is not simultaneously reusable, then $\mathcal{A}$ must hold concurrent float $\geq nP$. If, in addition, the estimator clusters $\mathcal{A}$-controlled identities—so that

evasion forces the interior nodes of each funding path to be coalition-external—then each path also carries at least $k-1$ nonrecoverable relay fees ϕ paid to those external nodes (Lemma 1), and $C_{\text{mimic}} \geq nP\,k\,\tau_{\text{fl}}\,r + n(k-1)\phi$.

Proof. By the source-coverage premise, every simulated requester has a funding path beginning at a source already recognized as $\mathcal{A}$-linked. Evading clause (i) forces each such path to exhibit at least k hops, and evading clause (ii) forces a rest of at least τ_{fl} before each hop, so the capital P earmarked to a path is immobilized for at least $k\tau_{\text{fl}}$ (transfer latencies only add). *Cost:* path i contributes committed capital-time $\geq P\,T_i \geq Pk\tau_{\text{fl}}$, and these are additive across the n distinct paths, so total committed capital-time is $\geq nPk\tau_{\text{fl}}$; capital cost is the rate r times committed capital-time, giving the first bound. *Concurrent float:* writing a_i for requester i's activation and d_i for the instant its P is first immobilized, $a_i - d_i \geq k\tau_{\text{fl}}$; with $a_{\max} - a_{\min} \leq \Omega < k\tau_{\text{fl}}$ one gets $d_i < a_{\min} \leq a_i$ for all i, so at $t^\star = a_{\min}$ every path has begun but not finished, and the n commitments draw on disjoint capital: concurrent float at $t^\star$ is $\geq nP$. *Fees:* if the estimator clusters $\mathcal{A}$-controlled identities, any interior node under $\mathcal{A}$'s control contracts the visible path below k hops and triggers the flag, so the $k-1$ interior nodes must be external; by Lemma 1 each is paid a relay fee ϕ that never returns to $\mathcal{A}$, adding $n(k-1)\phi$. □

Both quantities are read off the estimator rather than posited: k is its depth threshold and τ_{fl} its dwell threshold. At an annual capital-cost rate of 10% and total dwell $k\tau_{\text{fl}} = 30$ days, however, the interest term is only $0.10 \cdot 30/365 \approx 0.82\%$ of simulated nominal volume. This estimator-relative form is the contribution; dwell alone need not create an orders-of-magnitude floor. On a volume-weighted election rule the bound acts as an implicit tax on simulated volume at rate $k\tau_{\text{fl}}r$, in the same units as the explicit rate θ of Corollary 3. The comparison favors the explicit burn: a one-percent payment tax already exceeds this example estimator's floor, and it holds without source coverage and without any estimator. Source coverage remains essential. If the coalition acquires funds through an exchange, mixer, OTC counterparty, or another source not recognized as $\mathcal{A}$-linked, the proposition gives no bound for that capital; its acquisition cost must instead be included separately in C_{mimic}. Funding-graph and bridge-clustering analyses used against airdrop farming and wash trading [9, 10] motivate the estimator class, but their empirical coverage and classification error are not established by the proposition.

Remark 10 (Escape hatches). Costly identity, proof of personhood, KYC, hardware-bound identity [20], trusted demand oracles, and non-transferable credentials evade Theorem 1 by violating Assumption 1 or 2. Random protocol assignment of jobs to registered workers does not by itself escape: with free worker registration, a coalition fills the assignment pool with Sybil workers and recovers the payment when one is selected. If worker registration is resource-gated, the security comes from that gate and must be priced as a resource coordinate. This is the Douceur-shaped boundary [2]: each successful mechanism imports scarcity or trust from outside the receipt system, and the design question becomes *where* the imported root sits—not whether one is needed.

5 Collapse modes and scalar-cost equivalence

Corollary 1 (Two independent failure modes)**.** *For a demand-weighted mechanism:*

(i) ***Cheap-weight collapse.*** *If the minimum irrecoverable execution cost per unit of granted weight approaches zero—because the coalition can select or construct cheap instances—then by Lemma 2 leader election degenerates toward a fee-burn Sybil auction: weight is purchasable*

at approximately $F_{\text{burn}}(q) + C_{\text{capital}}(q) + C_{\text{mimic}}(q) + \mathbb{E}[S(q)] - V^{\mathcal{A}}_{\text{salvage}}(q)$*, with* $V^{\mathcal{A}}_{\text{salvage}} \approx 0$ *when the constructed cheap instances carry no salvage value.*

(ii) ***Usefulness collapse.*** *If every receipt requires expensive computation (*$\mathsf{W} = 1$*) but the computation is not bound to demanded jobs (*$\mathsf{B} = 0$*), consensus may remain computationally secure while the claimed usefulness evaporates: the work performed is real but serves no one.*

The two modes are orthogonal, and E overlays both: by Theorem 1, even $\mathsf{W} = \mathsf{B} = 1$ receipts cannot certify that their payments came from an independent demander—a *demand-exogeneity failure*, distinct from either collapse mode. The three properties are pairwise non-implied, and all six directions are witnessed: $\mathsf{W} \not\Rightarrow \mathsf{B}$ by usefulness collapse and $\mathsf{B} \not\Rightarrow \mathsf{W}$ by cheap authenticated service (the two off-diagonal cells of Table 1); $\mathsf{W} \not\Rightarrow \mathsf{E}$ and $\mathsf{B} \not\Rightarrow \mathsf{E}$ by Theorem 1, since a costly, well-bound receipt still cannot prove independent payment; and the two demand-antecedent directions by ordinary commerce—an economically independent customer can pay for a receipt that carries no work-cost floor ($\mathsf{E} \not\Rightarrow \mathsf{W}$: a cheap paid inference call) and whose payment is not bound to any fresh committed job ($\mathsf{E} \not\Rightarrow \mathsf{B}$: a prepaid retainer settled before the work is specified). Table 1 summarizes.

	Job binding absent (B=0)	**Job binding present** (B=1)
Irrecoverable cost floor (W=1)	*Usefulness collapse*: reported for Pearl cuPOW [22]	Useful work occurs; external subsidy unprovable without E (Thm. 1)
Weight below work cost (W=0)	Pure wash-weight mechanism	Authenticated service, insecure election; deployed analogue: incentive-program farming (§7)

Table 1: Failure taxonomy. No endogenous receipt mechanism simultaneously establishes a lower-bounded consensus cost and proves that nominal service payments came from an economically independent demander.

Corollary 2 (Scalar-cost equivalence, attacker-side)**.** *Define the comparison metric as the* attacker's minimal net irrecoverable cost per unit of election weight, with honest supply held fixed. *In a model where consensus resources are perfectly fungible, instantly purchasable, and security-relevant only through their net irrecoverable monetary cost, receipt-derived election weight provides, on this metric, no greater Sybil-resistance lower bound than an explicit fee burn equal to* $\max\{0, U_{\mathcal{A}}(q)\}$*. If receipt verification introduces positive overhead, the burn mechanism achieves a floor no lower on this metric with a strictly smaller verification surface—a two-coordinate dominance; the metric itself does not price overhead, so no single-metric dominance is asserted.*

Proof. In the scalar model with honest supply held fixed, consider the explicit-burn benchmark defined by the burn schedule $F(q) = \max\{0, U_{\mathcal{A}}(q)\}$ required to acquire weight q: its attacker minimal net irrecoverable cost for weight q is exactly $F(q)$, since the burn is nonrecoverable and fungibility with instant purchase eliminates every other cost channel. For the receipt mechanism, Lemma 2 gives $C^*_{\mathcal{A}}(q) \le U_{\mathcal{A}}(q) \le F(q)$ for every q, so the best valid attack-cost lower bound available to receipt-derived weight does not exceed the burn benchmark at any weight, and hence not per unit of weight either. Holding honest supply fixed is what disables the equilibrium channel of Proposition 2; that channel is priced separately (Remark 11, Corollary 3). The overhead clause is the two-coordinate statement above: floor no worse by the argument just given, verification surface strictly smaller by Definition 6. □

The burn schedule $F(q) = \max\{0, U_{\mathcal{A}}(q)\}$ is a comparison device. It is calibrated to the attacker's private costs, including salvage, so no protocol can set it; Corollary 2 compares the two mechanisms' surfaces at an equal floor and does not claim that a designer can implement that floor.

Remark 11 (Scope of the scalar metric). The ceteris-paribus clause matters: equal attacker cost per receipt does not imply equal *system* security, because the two mechanisms may attract different honest equilibrium supply. A receipt mechanism riding genuine service demand can accumulate a large demand-funded honest weight W_H, while an otherwise-equivalent isolated burn attracts only reward-motivated burners; by Proposition 2(iii) the former may be more secure in aggregate while the demand flow exceeds the threshold (4). Corollary 2 is therefore an attacker-side statement, and the fair system-level comparison is not receipt weighting versus an isolated burn but receipt weighting versus a *demand-funded* burn—which the next corollary constructs. The transient advantage of demand-funded honest weight is real, is captured by that construction, and is excluded from the security floor by Proposition 4.

Corollary 3 (Payment-tax equivalence: the demand-funded burn)**.** *Let a mechanism levy a tax at rate $\theta \in (0, 1]$ on every settled job payment,* burned outright*, and assign election weight in proportion to tax burned rather than to receipt count or nominal volume. Then: (i) self-dealt payments incur the same burn as genuine ones, so by Lemma 1 the coalition's cost per unit weight is bounded below by the burn, restoring a nonzero floor that nominal-volume weighting lacks; (ii) genuine demand funds honest weight through the same channel, capturing the demand-contingent (above-threshold, $v_{\mathcal{A}} \geq v_H$) subsidy of Proposition 2(iii) without any demand-authentication assumption—a* channel identification*, not a derived magnitude: how much of the subsidy reaches honest weight depends on tax incidence (payer, provider, or split) and on the participation margin—whether genuine volume settles through the taxed channel at all when untaxed venues compete; self-dealt volume participates by construction (the burn is what buys weight), while genuine volume participates only where settlement surplus plus reward pass-through covers its share of θ—both of which we leave unmodeled. Those margins determine the honest subsidy, not the self-dealer's burn floor; full exit of genuine demand is the projected zero-demand case of Proposition 4, whose funding equilibrium the proposition does not model. The floor of clause (i), by contrast, is incidence-invariant: for self-dealt volume the payer and the provider are the same coalition, so whoever nominally bears the tax, the coalition burns θ per unit of volume; and (iii) the mechanism's security resource is exactly the burn, so it inherits the scalar-model analysis of Corollary 2 with $F_{\text{burn}}(q) = \theta \cdot (\text{taxed volume})$, and requires no proof of E from any receipt. Demand-weighted election secured by a payment tax is thus an explicit burn in receipt clothing—which is precisely why it works: the security claim rests on the nonrecoverable component, never on demand exogeneity.*

Three boundary conditions are part of the statement. First, the taxing chain must observe final settlement before the burn enters election: either it consumes receipts finalized in a strictly earlier epoch, or an independently finalized settlement channel itself implements and proves the burn. Second, the taxed portion must be destroyed, not redistributed: tax flow routed to validators, stakers, or a treasury is recapturable by coalition members in those roles, and Lemma 1 then voids the floor of clause (i). Third, the burn branch is the only endogenously sound one: a variant paying the tax to "verifiably coalition-external" parties would require certifying coalition-externality of the payee—precisely what Assumption 2 and Theorem 1 make endogenously uncertifiable—and is therefore available only under an imported trust root (Remark 10). Burning as a consensus resource has cryptographic precedent [30]; the corollary's content is the reduction of demand-funded election weight to it.

Burning the native asset is, to first order under aggregate-value invariance, a pro-rata rebate to remaining holders: a coalition holding float share η recovers on the order of η of any burn through appreciation, so its effective clause-(i) floor is about $(1-\eta)\theta$ per unit taxed volume. The rebate is holder-symmetric—honest holders receive the same per-unit subsidy, so it enters Proposition 2 as symmetric salvage, degrading security only insofar as the coalition's float share exceeds honest burners'—and it is fork-dependent, denominated in the asset a safety attack endangers, hence outside $V^{\mathcal{A}}_{\text{salvage}}$'s fork-independent accounting for consensus-breaking attacks (though fully operative in the centralization mode of Remark 6, where no attack on token value is intended).

Genuine and self-dealt volume respond differently to the tax. A genuine payer may move to an untaxed venue or decline to transact as θ rises, while a self-dealer treats θ as the direct price of election weight. Raising θ can therefore shrink honest burn faster than adversarial burn and worsen the composition of taxed volume. An interior optimum may exist; locating it requires demand elasticities and reward pass-through outside this framework.

Remark 12 (Burden shift, not universal dominance). Corollary 2 is a scalar-model statement, not an unconditional impossibility. Compute is time-bound where capital is instant; hardware and capital have different market depth; multi-resource requirements may resist concentration; useful demand can subsidize honest participation (Proposition 2(iii), priced by Corollary 3); physical work may rate-limit weight acquisition. A hybrid mechanism may legitimately claim security from non-substitutable resources—but the claim must then be *modeled*: any advantage of receipt weighting over an equivalent burn must arise from a separately proved property (rate limits, non-fungibility, honest-side subsidy, distributional effects), never from nominal job payments or receipt volume, which Lemma 1 voids. Proposals that weight election by receipt count or gross service value without such a model inherit the corollary.

6 The demand-independent security floor

The architectural robustness lemma in this section does not depend on demand indistinguishability (Theorem 1) or on the salvage model (§3). Premise P4 is an explicit design norm, not a consequence of permissionlessness; demand-dependent consensus is the stated escape. Premise P5 supplies the projection closure needed by the proof and prevents the null-demand restriction from creating an execution that was infeasible in the original mechanism.

We take safety and liveness in the standard sense, against a Byzantine adversary that may deviate arbitrarily under the mechanism's network model: *safety* is the absence of conflicting finalized histories, and *liveness* is eventual progress. Proposition 4 inherits the original mechanism's synchrony, corruption-timing, and adaptive-security assumptions, with one explicit modification: adversarial admissibility is quantified over resource coordinates, not over total election weight after a demand boost. The threat model bounds the coalition's resource profile—its share of w_r from attested nonrecoverable work, stake, burn, or lockup—and imposes no separate bound on self-authored demand coordinates. P4 must therefore cover an admissible resource-bounded coalition even when realizable demand inflates its total election share past the protocol's safety threshold; a mechanism that fails on that execution violates P4 rather than excluding the execution after the fact. The share threshold $\tau \leq \frac{1}{2}$ is needed only for Corollary 4, not for the projection argument itself.

Proposition 4 (Demand-independent security floor)**.** *Consider a mechanism, in the sense of Definition 2, required to satisfy:*

P1. *permissionless participation (Assumptions 1–2);*

P2. *customer-authorized workloads: jobs are selected and authorized by requesters; the protocol originates no jobs and no instance schedules;*

P3. *nonrecoverable-weight accounting: the mechanism's security argument counts only value nonrecoverable by the weight-holding coalition (Lemma 1); this premise scopes the mechanism class and is not invoked in the proof below—it is consumed by Corollary 6 and the design conclusions of §8, and is stated here so that P1–P5 characterize the class in one place;*

P4. *consensus safety and liveness in every admissible demand environment of §2, including the* zero-demand path*: zero external job demand from genesis onward. Coalition admissibility is fixed by the resource-coordinate threat model above and is not withdrawn when self-authored demand raises the coalition's total election share;*

P5. *no protocol-originated demand and null-demand projection closure: the protocol neither manufactures filler jobs, schedules customer instances, nor prescribes payments on its own authority. For every feasible zero-demand execution of π_r, the same resource actions and adversarial/network behavior are feasible in the original mechanism with each demand coordinate set to its canonical null value x_d^0. The protocol may prescribe and attest* work *in zero-price jobs (Remark 13), but not require participants to pay one another.*

Then demand-derived receipt weight *(Definition 4) cannot be a necessary leader-election resource, necessity in the sense of Definition 5: the mechanism must possess an election path whose safety and liveness do not depend on the demand-derived component of receipt weight. The claim is existential. Its scope—silence about dependence confined to demand-rich environments, and about receipts whose election contribution loads only on attested work or burn content—is discussed after the proof (Remark 13).*

Proof. Take an arbitrary admissible zero-demand execution E_r of the restricted mechanism π_r, including arbitrary adversarial behavior allowed by the mechanism's threat model. By P5, there is a corresponding execution E^0 of the original mechanism with the same resource actions and network/adversarial behavior, but with every demand coordinate set to its canonical null value x_d^0.

By Definition 4, every identity has weight

$$w(x_d^0, x_r) = w_r(x_r)$$

in E^0. The election weights and election-relevant execution of E^0 are therefore identical to those of E_r. Premise P4 guarantees safety and liveness for E^0, because it is an admissible zero-demand execution of the original mechanism. The identical election execution E_r is consequently safe and live as well.

Since E_r was arbitrary, no admissible zero-demand execution fails under π_r. Demand-derived receipt weight is therefore not necessary in the sense of Definition 5, and π_r is the required election path. Notice what this proof does not assert: it does not call every variation in $w - w_r$ costless. Resource-coupled boosts are covered by the null-demand projection, while free realizability is an additional premise used only by Corollary 4. □

The proposition establishes feasibility of the projected path, not an economic supply equilibrium for it. Sizing follows only from the resource-quantified reading of P4 stated above and is made explicit in Corollary 4. Whether honest participants have revenue with which to supply the demand-free resources when external demand is zero is a separate funding question not modeled here; it is the zero-demand analogue of the fee-only instability studied by Carlsten et al. [5].

Mechanisms that accumulate demand-derived weight while demand lasts and then coast on that backlog fall outside the class, because P4 quantifies the zero-demand path from genesis (Remark 14); the proposition says nothing about them.

Remark 13 (Work in receipt syntax). The qualifier *demand component* is necessary: the following construction shows why the stronger claim that "receipts cannot be the election resource" would be false. Let the honest protocol prescribe that every participant self-author, execute, and settle one *zero-price* job per epoch (program and inputs of its choice, fully contractible acceptance, payment terms fixed to zero), with election weight equal to the attested irrecoverable work in receipts valid per (1) and finalized in earlier epochs. The settlement clause Settle of (1) is still discharged—it finalizes a zero payment—so the receipts are valid, yet no payment changes hands and the weight loads only on attested work. This mechanism satisfies P1–P5—in particular P5, because the prescribed jobs carry zero payment terms, so the protocol prescribes *work* and a trivial (payment-free) settlement, never a payment—and stays live on the zero-demand path, with receipts its only election resource. Its P4-safety imports the standard assumption of every work-weighted consensus, here made explicit: honest parties hold a super-threshold share ($> 1 - \tau$) of attested work; and first-epoch election is seeded by a genesis rule (a designated genesis allocation, or an initial bootstrap epoch that finalizes no transactions until attested work exists), after which epoch-k election reads work finalized in epochs before k. It is no counterexample to the proposition: all of its election weight is *work-derived*; the payment content of its receipts is recoverable and carries nothing. Such a design is proof of work in receipt syntax, and its floor is exactly its nonrecoverable work content—an instance of the proposition's conclusion, not a refutation. The variant that instead prescribes self-*payment* and weights election by payment content is the case P5 excludes: it manufactures demand-derived content on the zero-demand path, and its "floor" is a recoverable transfer (Lemma 1), which is exactly the failure the proposition asserts. The proposition excludes a floor constituted by the *demand* content of receipts—payment volume, receipt count, gross service value, or any weight loading on requester-side content—whether that content arises from real demand or from protocol-prescribed self-payment.

Remark 14 (Genesis quantification and backlogs). The zero-demand path is quantified from genesis deliberately: a weaker premise—demand that merely *becomes* zero—admits mechanisms that stockpile demand-derived weight during a demand era and coast on the backlog, evading the argument while remaining demand-dependent. Quantifying from genesis closes that evasion but strengthens P4 and therefore narrows the mechanism class to which the proposition applies. This is a robustness bound, not a prediction: the claim is not that demand will be zero, but that a mechanism unable to survive that path has a demand-contingent floor. External demand may genuinely raise security while it exists; it cannot supply the component required to survive its disappearance.

For nonnegative boosts satisfying Definition 4(iv), P4 yields a quantitative companion. The bound is profile-dependent and therefore applies to resource-coupled boosts as well as uniform ones.

Corollary 4 (Sizing rule for the receipt boost)**.** *Fix an in-class mechanism (P1–P5) with adversary-share safety threshold $\tau \le \frac{1}{2}$ and the resource-coordinate adversary semantics stated above. At an admissible resource profile (x_r^A, x_r^H), write $\omega_A := w_r(x_r^A) \ge 0$ and $\omega_H := w_r(x_r^H) > 0$. Let $\beta_H \ge 0$ be the honest demand-derived boost, and let $b_A \ge 0$ be any boost the coalition can realize at fixed x_r^A through self-authored demand at nonrecoverable cost bounded independently of the requested*

boost. P4 safety on the corresponding flood execution requires

$$b_A \;<\; \frac{\tau\,(\omega_H+\beta_H)-(1-\tau)\,\omega_A}{1-\tau}, \qquad \textit{equivalently} \qquad \omega_A+b_A<\frac{\tau}{1-\tau}(\omega_H+\beta_H).$$

Consequently, the profile-dependent supremum $B_A(x_r^A) := \sup\{b_A : b_A \text{ is freely realizable at } x_r^A\}$ is at most the right-hand side, with strict inequality when the maximum is attained. The conservative case is $\beta_H = 0$. If instead the mechanism offers an attainable, uniformly freely realizable per-coalition cap $\beta^{\mathrm{f}}_{\max}$, that cap must satisfy the strict first inequality with $b_A = \beta^{\mathrm{f}}_{\max}$ for every admissible resource profile and therefore lie below the infimum of its right-hand side. If honest demand-free weight is not bounded away from zero over those profiles, no positive uniform cap is justified; a nonzero uniform boost requires an enforceable demand-independent participation floor. The conclusion is per coalition and does not bound aggregate demand-derived weight on benign executions.

Proof. On the flood execution the adversary holds $\omega_A + b_A$ and the honest set holds $\omega_H + \beta_H$. Because coalition admissibility was fixed before the boost by its resource coordinates, P4 applies to this execution and requires

$$\frac{\omega_A+b_A}{\omega_A+b_A+\omega_H+\beta_H}<\tau.$$

Multiplication by the positive denominator gives $(1-\tau)(\omega_A+b_A)<\tau(\omega_H+\beta_H)$, hence the display. Taking the supremum gives the weak bound for B_A; attainment gives the strict form. For the resource-coupled example $w = W(1+\min\{V,\kappa W\})$, the maximum $B_A(W_A)=\kappa W_A^2$ is attained, so at $\tau=\frac{1}{2}$ the obligation is $W_A(1+\kappa W_A)<W_H+\beta_H$. Thus $W_H=100$, $W_A=10$, $\kappa=1$, and $\beta_H=0$ is unsafe: self-paid $V=10$ gives adversarial weight 110 and share $110/210>1/2$. Resource coupling changes the bound from a constant to a generally nonlinear function of the adversary's resources; it is not an exemption from sizing. □

Corollary 4 covers boosts a coalition can realize for free. A boost priced by a burn, such as the payment tax of Corollary 3, is not free, and Definition 4(i) classifies its weight as burn-derived, so Proposition 4 does not exclude it. The difficulty sits in the honest-majority premise instead. When demand vanishes, honest parties stop burning, while a coalition can keep its own taxed volume alive by paying itself at rate θ; the protocol sees the same taxed volume either way (Theorem 1). A premise stated over stake plus burn therefore becomes, at zero demand, a premise over stake that the attacker can top up. The next corollary states the safety condition in terms of the floor resource, the quantity a designer can reason about when demand is gone. The burn sets what the attack costs; the cap sets whether it succeeds.

Corollary 5 (Sizing for burn-priced boosts)**.** *Let election weight be $w_i = s_i + g_i$, where $s_i \ge 0$ is identity i's demand-independent floor resource and $g_i \ge 0$ is a boost acquired by burning value at any positive price, with adversary-share safety threshold $\tau \le \frac{1}{2}$. Write $\omega_A, \omega_H > 0$ for the coalition's and the honest set's floor weight and $a = \omega_A/(\omega_A+\omega_H)$ for the coalition's floor share. The threat model bounds a and places no bound on how much the coalition chooses to burn. On an execution where honest demand, and hence honest boost, is zero:*

- *(i)* ***Stake-coupled cap.*** *If $g_i \le \kappa s_i$ for every identity, safety requires $a < \tau/(1+\kappa(1-\tau))$; at $\tau=\frac{1}{2}$, $a < 1/(2+\kappa)$.*
- *(ii)* ***Aggregate cap.*** *If only aggregate boost is capped, $\sum_i g_i \le \kappa \sum_i s_i$, safety requires $a < \tau - \kappa(1-\tau)$; at $\tau=\frac{1}{2}$, $a<(1-\kappa)/2$, and no positive floor share is safe once $\kappa \ge \tau/(1-\tau)$.*

(iii) ***No cap.*** *If boost weight is unbounded, no floor share $a > 0$ is safe.*

None of these bounds depends on the burn price, the tax rate θ, or whether burned weight decays; those determine the cost of the attack. The cap in (i) is Sybil-proof because splitting stake across identities leaves $\sum_i \kappa s_i$ unchanged. A per-identity cap that is not proportional to a committed resource bounds nothing under Assumption 1.

Proof. With honest boost zero, the coalition's share $(\omega_A + g_A)/(\omega_A + g_A + \omega_H)$ is increasing in its boost g_A, so safety requires $(1-\tau)(\omega_A + g_A^{\max}) < \tau\,\omega_H$ at the largest boost available to it. (i) Summing the per-identity caps gives $g_A^{\max} = \kappa\omega_A$; dividing $(1-\tau)(1+\kappa)\omega_A < \tau\omega_H$ by $\omega_A + \omega_H$ gives $(1-\tau)(1+\kappa)a < \tau(1-a)$, that is, $a < \tau/(1+\kappa(1-\tau))$. (ii) With honest boost zero the coalition can take the whole aggregate cap, $g_A^{\max} = \kappa(\omega_A + \omega_H)$; dividing gives $(1-\tau)(a+\kappa) < \tau(1-a)$, that is, $a < \tau - \kappa(1-\tau)$. (iii) Without a cap the left-hand side is unbounded for every $\omega_A > 0$. □

Part (i) is the burn-priced counterpart of the resource-coupled example in Definition 4(ii). Coupling the boost to a committed resource turns an unsafe uncapped boost into a bounded loss of tolerance, from τ to $\tau/(1+\kappa(1-\tau))$: with $\kappa = 1$ and $\tau = \frac{1}{2}$, a coalition needs one third of the floor rather than one half. The attestation-native design of Stefanović [24] has the form of (i). Its vote weight is bonded stake times a non-transferable reputation clipped to $[r_{\min}, r_{\max}]$ and earned through fee-weighted processing of attestations, with $r_{\max}/r_{\min} \in [4, 10]$ recommended and safety inherited from BFT under Byzantine weight below one third of the total. Taking stake times $r_{\min}$ as the floor resource gives our $\kappa = r_{\max}/r_{\min} - 1 \in [3, 9]$, not to be confused with that paper's cost-premium κ. With $\tau = \frac{1}{3}$ and honest reputation at $r_{\min}$, part (i) gives $a < 1/(3+2\kappa)$: a coalition that has raised its own reputation to $r_{\max}$ needs between about 5% ($\kappa = 9$) and 11% ($\kappa = 3$) of bonded stake, against one third under pure stake. Honest reputation sits at $r_{\min}$ on the zero-demand path from genesis, and for honest validators who join after demand ends, since re-entry resets reputation to $r_{\min}$. That paper analyzes the complementary case, a fresh attacker entering at $r_{\min}$ against honest validators of mean reputation $\bar{r}_H$, which is the source of its $\bar{r}_H/r_{\min}$ cost premium over pure stake. Its fee-burn bound prices the climb to $r_{\max}$ but, by Corollary 5, does not change the safe share. The design's safety margin therefore depends on honest reputation staying high, and reputation earned from fee-weighted attestation processing ties that to demand; the paper itself notes that its slash deterrent attenuates in low-volume periods. Windowing burned weight changes the cost of building a boost in advance (the backlog of Remark 14) but not these bounds, which already assume the coalition holds the full cap.

Proposition 4 does not by itself forbid *hybrid* designs that add bounded receipt-derived weight on top of the demand-independent floor. Full separation is a design conclusion within an explicit resource model, which we now state.

Definition 6 (Weak dominance and scalar-model minimality (P6))**.** We measure a weight component by two sets. Its *verification surface* is the set of predicates a validator must evaluate inside consensus per unit of weight granted. For an explicit burn, a validator checks whether a finalized transaction assigned amount b to a provably unspendable output. For a receipt, it checks the full relation (1). Because $\mathsf{Settle}(s, J, a)$ already verifies a finalized transaction assigning a specified amount to a specified payee, the burn predicate is the instance of Settle whose payee is the unspendable output; the receipt relation additionally carries Auth, Fresh, Exec, $\mathsf{Deliver}$, and Accept_t. The containment is therefore strict for receipt-shaped work.

The *manipulation surface* is the set of attacker-controllable inputs in those predicates. Under our canonicalization, a burn exposes only its amount; a receipt exposes payment amount, job choice,

timing, provenance, and acceptance. The former is contained in the latter, with payment amount as the receipt analogue of burned amount. A component X *weakly dominates* Y when X provides at least Y's quantified floor and, after this canonicalization, X's verification and manipulation surfaces are subsets of Y's, at least one strictly. Dominance is a partial order. In particular, P6 does not dominate hash-puzzle proof of work: its verification surface is incomparable to a burn's, not a superset. P6 targets receipt-shaped work because receipts add execution, delivery, and acceptance predicates to settlement. Premise P6 (*scalar-model minimality*): within the scalar-cost model of Corollary 2, the mechanism excludes any auxiliary election component—demand- or work/burn-derived—that is weakly dominated by an explicit nonrecoverable-cost mechanism providing the same quantified security floor. Components with incomparable surfaces fall instead to the quantification burden of Remark 15.

Corollary 6 (Scalar-model settlement–election separation)**.** *Under P1–P5, the assumptions of Corollary 2, and P6 (Definition 6), service receipts carry no leader-election weight: settlement and election are fully separated. This is a design conclusion within the scalar benchmark model, not an impossibility. Note its scope is broader than Proposition 4's: the proposition removes only the* demand-derived *component from the necessary floor, whereas here P6 plus the scalar premise remove* all *receipt-derived election weight, demand- and work/burn-derived alike.*

Proof. By Proposition 4, the mechanism has an election path independent of demand-derived receipt weight that is safe and live on the zero-demand path; this secures safety and liveness without any demand-derived component under P4. It remains to remove work/burn-derived receipt weight as well, which is where P6 and the scalar model do the work. Any auxiliary receipt-derived weight—of either provenance—is acquirable by a coalition at cost $C^*_{\mathcal{A}}(q) \le U_{\mathcal{A}}(q)$ (2), in which nominal payments do not appear (Lemma 1). Within the scalar-cost model it establishes no security floor beyond an explicit burn of $\max\{0, U_{\mathcal{A}}(q)\}$ (Corollary 2); yet, by Definition 6, it carries the full validity relation (1) as verification surface and job/timing/provenance/acceptance freedom as manipulation surface, each a proper superset of the burn's surfaces, and when the work has salvage value it also weakly lowers the coalition's unit cost of weight (Proposition 1), with the security consequence governed by the regimes of Proposition 2. The explicit burn thus weakly dominates the receipt-derived component—equal floor, strictly smaller surfaces—so P6 excludes it in favor of the explicit-burn form (demand-funded where service activity is to contribute, Corollary 3). □

Remark 15 (Quantified non-scalar hybrids). Mechanisms claiming a genuinely non-scalar benefit from receipt weighting—rate limits, resource non-substitutability, honest-side subsidy (Remark 12)—fall outside Corollary 6, whose premise is the scalar model. They do not refute Proposition 4: their safety and liveness still require a floor independent of demand-derived receipt weight, and their claimed incremental security must be established separately, with the receipt-derived component's cost accounted net of recoverable transfers and salvage.

Remark 16 (The strongest compliant hybrid). The strongest design permitted by our results combines an unconditional proof-of-stake floor sized to carry safety and liveness alone; a receipt-derived election boost subject to an explicit design cap; an irrecoverable tax on every receipt payment (Corollary 3) so that the boost's security contribution is an explicit burn; provider bonds with randomized audits converting service fraud into collectible slashing ($\mathbb{E}[S] > 0$); optionally, externally authenticated demand for the boost's eligibility (stepping outside Assumption 2 where its operator accepts that trust root); and a window after which burned weight expires, so that the honest boost decays when genuine taxed demand vanishes. Decay cannot remove a coalition's

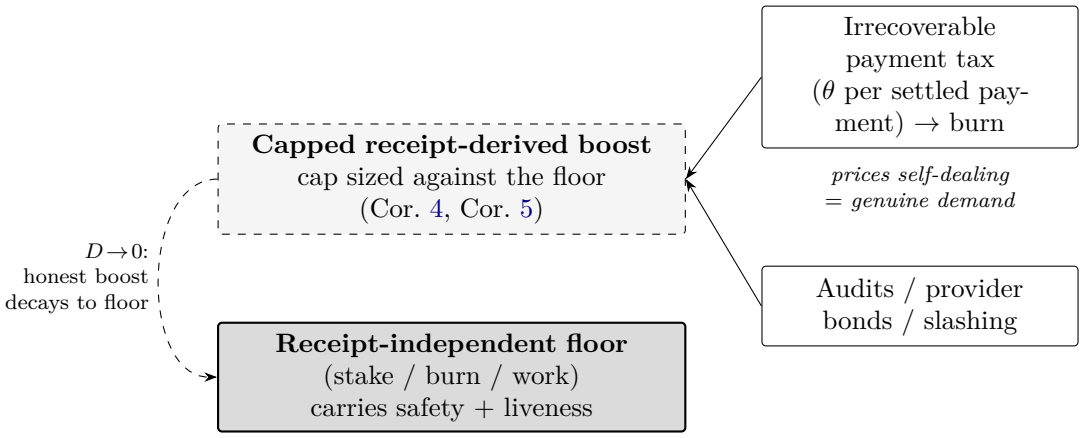


Figure 2: The floor-plus-capped-boost architecture (Remark 16, §8). An unconditional, receipt-independent floor (stake / burn / work) carries safety and liveness on its own, including the zero-demand path. A receipt-derived boost sits above it under a cap sized against the floor alone (Corollary 4 for freely realizable boosts, Corollary 5 for burn-priced ones). Its security rests not on nominal demand but on an irrecoverable payment tax burned outright (so self-dealing is priced exactly like genuine demand, Corollary 3) and on audit-backed provider bonds with collectible slashing. As genuine taxed demand vanishes ($D \to 0$) the honest boost decays to the floor. A coalition can keep its own boost alive by paying the burn, which is why the cap, not the burn price, carries safety.

boost, because the coalition can keep its own taxed volume alive by paying the burn. If any part of the boost remains freely realizable, that part must satisfy Corollary 4; a burn-priced part is not free, but its cap must still be sized against the floor alone (Corollary 5). Every element is consistent with P1–P5 (Proposition 4) and P6 (Definition 6) as applicable: the stake floor is the demand-independent path of Proposition 4; the boost's security rests on its burn and slashing components, not on nominal demand; and on the zero-demand path the honest side falls back to the floor while any coalition's boost stays within the cap. The framework permits this design. Filecoin has run a floor-plus-attested-boost architecture in production since 2020 (§7), and a recent attestation-native consensus proposal [24] independently instantiates its core—validator vote weight is bonded stake times a non-transferable reputation clipped to $[r_{\min}, r_{\max}]$ and earned through fee-weighted attestation processing, with a cost-to-grind floor priced by an adaptive fee-burn fraction and "inherited BFT safety and liveness"—a receipt-independent floor plus a bounded, burn-priced receipt component. The framework supplies the accounting discipline: no endogenous receipt proves economically independent demand (Theorem 1); any consensus security attributed to receipts must come from nonrecoverable cost or a separately quantified non-scalar property (Corollaries 2–3, Remark 15); and safety and liveness require a floor independent of demand-derived receipt weight (Proposition 4). This is narrower than claiming that receipts cannot secure a blockchain.

Remark 17 (Every alternative abandons a premise). The characterization is tight in the sense that each known escape drops exactly one premise or concedes the conclusion: protocol-scheduled instance pools, where the protocol itself originates the epoch's problems, abandon P2 or P5; designs that decouple consensus hardness from client demand—Ofelimos [13] mines by stochastic local search with a miner-generated, hash-derived fallback when client-posted instances are absent—retain a demand-independent work floor, conceding the path of Proposition 4 by construction;

trusted demand oracles and KYC abandon P1; hardware attestation [20] imports a vendor trust root, abandoning Assumption 2; fee-burn election abandons work-derived weight altogether; demand-dependent consensus abandons P4; hybrid boosts claiming non-scalar benefits step outside Corollary 6's scalar premise and inherit its burden of quantification; and fallback PoW/PoS "when demand is low" concedes the demand-independent path of Proposition 4—which is the floor, implemented reluctantly. Among mechanisms preserving permissionless customer authorization while refusing to count recoverable payments as Sybil-resistant cost, separating settlement from election is the canonical architecture: self-dealing then buys no consensus weight, and a coalition paying itself for useless service has merely performed an economically pointless transaction.

7 Deployed evidence

Pearl cuPOW: usefulness collapse as a design property. Pearl combines a matrix-multiplication work certificate with the claim of useful AI inference. Basu's preprint audit [22]—single-author and, at the time of writing, an unreviewed preprint—reports that the most widely used publicly released miner contains no inference path, that the verification protocol accepts adversarially chosen operands including uniformly random matrices, and that a purpose-built miner performing no inference earned accepted shares across four hardware families. The audit reports network-wide scale (24 EH/s and 8,012 analyzed workers on a pool at roughly 21% of network hash rate) but does not attribute a network hash-rate share to the particular released miner binaries it inspects. We therefore scope the behavioral finding to the audited mining paths, while the design-level fact—random operands accepted, hence $\mathsf{B} = 0$—is protocol-level and independent of usage share.

The finding concerns a design property rather than a patchable vulnerability. The cuPOW verification relation establishes that a matrix product was computed correctly; it never purported to establish that the operands were anyone's data. Hardness is guaranteed, while utility is not—a gap the underlying construction acknowledges and the earlier property literature named theoretically [17, 18]. Basu's audit contributes a measurement: under ordinary profit-seeking, the useful work the network was marketed on did not occur in the measured mining paths.

In our vocabulary, and taking the report at face value: $\mathsf{W} = 1$ *under cuPOW's assumed computational cost floor* [21] (real GEMM work with a conjectured irrecoverable cost), $\mathsf{B} = 0$ (no binding to any demanded job), $\mathsf{E} = 0$ (no external subsidy), and $V^{\mathcal{A}}_{\text{salvage}} \approx 0$ (random operands have no salvage value—the discount of Proposition 1 is inactive precisely because the work is useless). Consensus security therefore did *not* collapse to fees; the work remained costly. What collapsed was usefulness. Pearl is thus the deployed instance of Corollary 1(ii), and it isolates the paper's central claim in the simplest possible setting: a protocol that verifies arithmetic rather than demanded jobs gives miners no reason to supply anything but arithmetic, and in the audited paths they did not. No strengthening of the execution proof repairs this, because the missing predicate is B, not Exec. Pearl reaches the $\mathsf{B} = 0$ column without ever encountering the E question; §4 shows that supplying B would not have sufficed either.

Incentive-program farming: a deployed analogue. A structurally identical failure—though against reward allocation rather than leader election, and hence an *analogue* rather than a second instance—is deployed at scale wherever protocol rewards are weighted by service volume: airdrop and points programs. LayerZero's 2024 response evaluated roughly six million interacting addresses through self-report and community-bounty phases [38]; the official post does not establish a final adjudicated Sybil count, so we do not use one here. Hop DAO's sybil-hunter process excluded over

ten thousand farmed addresses and returned about 3.5M HOP to its treasury [39]. A reward-side analogue is deployed in crypto-AI specifically: Bittensor's Yuma Consensus distributes token *emissions* using validator-assigned, stake-weighted miner scores; its own analysis of weight copying and commit–reveal defenses [35] shows that even the reward layer must address metric mimicry. We make no claim here about Bittensor's base-chain proposer mechanism: the analogy is reward-side only. Farming coalitions submit fully authorized, fresh, executed, delivered, and settled transactions to themselves—$\mathsf{B} = 1$ in every cryptographic respect—while $\mathsf{E} = 0$: the demand is self-dealt and the internal payments are recoverable. The mechanism of failure is Lemma 1 plus Lemma 2—provenance-blind volume weighting is self-dealt at zero net cost—with Theorem 1 supplying the defeat of provenance-*sensitive* defenses, and the defensive arms race—provenance clustering and funding-graph analysis, documented quantitatively for centralized- and decentralized-exchange wash trading [9, 10]—is exactly the estimator of Theorem 1(b). One disanalogy sharpens rather than weakens the lesson: an airdrop can be *retroactive*, announcing its qualifying criteria only after the qualifying window has closed, so that behavior matching the rule is harder to fabricate in anticipation—though not impossible, since criteria are often predictable from precedent. Consensus rules have no access to this defense—a leader-election rule must be prospective and public, and Theorem 1's simulator targets exactly the announced rule. Retroactivity is thus an escape hatch available to reward allocation but structurally unavailable to election weighting, one more asymmetry between what receipts can safely govern (settlement, rewards) and what they cannot (election). We know of no peer-reviewed measurement of *service-receipt* farming specifically, and flag this as an open empirical question; the general mechanism, however, is well documented. We are not aware of a production chain that weights leader election by *endogenously* attested demand under Assumption 2. The closest deployed system, Filecoin, multiplies election weight by demand attestation imported from a human notary root (next paragraph). Helium's original design provides an adjacent historical case in which proof-of-coverage scores influenced election into a consensus group [36]; HIP-70 later moved proof-of-coverage accounting to oracles and migrated the network to Solana for stated scaling and reliability reasons [37]. We therefore do not infer gaming as the cause of that migration. The *endogenous* election form of the failure remains a prediction of the model rather than a measurement; the reward-allocation analogues above cover the reward-side E analogue of Table 1, not its endogenous election-side center. The prediction is falsifiable: a production chain that weights leader election by endogenously attested demand at scale and sustains safety and liveness through a demand collapse without a receipt-independent floor would disconfirm not Proposition 4 but the assumptions under which it applies, and we would count it as evidence against the framework's applicability.

Gridcoin: deployed settlement–election separation. Gridcoin is a direct deployed precedent for the architecture recommended here. BOINC research affects research rewards but not the probability of producing a block: the project's documentation states that staking odds depend only on stake, while research changes only the reward [16]. It also states the zero-demand counterfactual explicitly: if every BOINC project shut down, proof-of-stake consensus would continue producing blocks and processing transactions, minus BOINC rewards. Gridcoin therefore separates useful-work reward accounting from its consensus floor rather than asking research receipts to provide that floor. This does not prove our impossibility results, but it narrows the novelty claim and shows that the design conclusion predates this paper in deployed form.

Filecoin: an attested-demand multiplier in production, and its floor. Filecoin is the deployed system closest to demand-weighted election, and it instantiates our conclusions rather

than refuting them. Expected Consensus elects block producers by a lottery over *quality-adjusted power* (QAP) [31]; committed-capacity sectors plus initial pledge keep raw-byte power and leader election fully operational at zero storage demand—a deployed demand-independent floor of exactly the shape Proposition 4 requires. On top of that floor the Fil+ program applies a 10× QAP multiplier to verified-deal data [32]: a demand-derived boost. But the protocol does not observe the demand—verified status is asserted by human *notaries* spending DataCap under a root-key multisig [33], so the boost's independence rests on an imported trust root (the escape of Remark 10). A governance incident illustrates the resulting trust-root risk: an official removal recommendation alleged DataCap awards without due diligence and collusion [34]. It is evidence about the attestation process, not a measurement of a consensus attack. Two of Filecoin's mitigations are exactly those our framework recommends: pledge collateral and penalties scale with QAP, so the boost is partly re-gated by a committed resource (Definition 4(ii)), and the floor carries safety independently of the boost.

HadAgent: applying the checklist to a live proposal. HadAgent [23] is a contemporaneous proof-of-inference proposal near our target class; we read its published design through the checklist, hedging where the preprint is underspecified. It ties block-creation rights to verified deterministic inference rather than to payment value or economically independent demand. That places the proposed resource on the *work-derived* side of Definition 4, where Proposition 4 permits it to be a zero-demand floor. The decisive question is therefore not whether a coalition can originate inference tasks, but what it must irrecoverably spend per unit of fresh, election-relevant inference weight. If the proofs are fresh, replay-protected, and impose the claimed compute cost, HadAgent is proof of work in receipt syntax (Remark 13), not the free demand flood of Corollary 4. If old proofs can be replayed or accepted tasks admit materially cheaper paths, the relevant failure is instead work soundness or cheap-weight collapse (Corollary 1(i)). The current preprint does not settle that audit: it invokes an honest majority of master nodes and leaves the selection, corruption cost, and proposer-cost parameters insufficiently specified for a security reduction. We therefore do not prescribe stake as a consequence of Proposition 4 or claim that HadAgent instantiates Proposition 4's excluded demand-derived floor. We record the narrower open obligation: specify the master-node trust and selection model, freshness/replay rules, the inference-cost lower bound, and how verified work maps to proposer probability.

8 Consequences for system design

Receipts as settlement instruments. A receipt satisfying (1) is strong evidence for *settlement*: it proves that an authorized, fresh job was executed, delivered, and accepted under its designated process, and it identifies the collateral answerable for failure. It cannot prove that the payment behind it represents independent demand. Receipts should therefore allocate payment, collateral, and losses; leader election should be secured by resources whose cost to an attacker is enforced outside the receipt system—externally scarce capital subject to lockup and collectible slashing, or explicit burns—rather than by demand artifacts a coalition can synthesize.

If demand must fund security, tax it. Corollary 3 converts the impossibility into a recipe. A designer who wants service activity to strengthen consensus should not weight election by receipts and hope the demand is genuine; they should levy a tax on every settled payment, *burn it outright* (destroyed, not redistributed—Corollary 3's boundary conditions), and weight election by tax burned. Self-dealing then buys weight only at the burn rate, genuine demand funds honest

weight through the same channel, and no component of the security argument rests on E. The residual design freedom is the rate θ and the incidence (payer, provider, or split); the residual obligation is Proposition 4: the taxed flow funds weight *above* the floor, never the floor itself. On the zero-demand path honest taxed volume is zero while a coalition's need not be, so the cap on tax-burned weight must be sized against the floor alone (Corollary 5). Tax-burned weight is burn-derived by the precedence rule of §2, so it lies outside the demand-derived component Proposition 4 excludes and is compliant by construction.

The centralization mode is not a Sybil mode. Remark 6 identifies a failure requiring no dishonesty at all: under demand-weighted election, a provider that wins the service market accumulates consensus power as a byproduct of being paid, and its net cost of weight can plausibly run negative (Remark 6). Defenses against wash demand—provenance estimation, funding-graph analysis, mimicry pricing—are inert against it, because there is nothing fake to detect. The only structural defenses are the ones this paper argues for on independent grounds: capping receipt-derived weight and carrying safety on a demand-independent floor. A designer evaluating a demand-weighted proposal should therefore ask two separate questions: *what does a coalition pay for synthetic demand?* (§3–4) and *what does the market leader get for free?* (Remark 6)—and note that the second question survives even a perfect answer to the first.

These design conclusions inherit their models, and the qualifier belongs at the point of use: full settlement–election separation is the scalar-benchmark conclusion (Corollary 6); for the non-scalar resource markets deployed systems actually occupy, the operative deliverable is the burden shift of Remark 12—any advantage claimed for receipt weighting over an explicit demand-funded burn must be separately proved, never presumed from volume.

The recursion onto service-layer reputation. Any settlement layer that scores providers must apply Theorem 1 to itself. Receipt count, nominal job value, gross volume, and unchallenged success streaks are all synthesizable at the coalition floor and must carry no reputational weight; this is the receipt-system instance of the impossibility of symmetric Sybil-proof reputation [41] and of the limits on Sybil-proof accounting of contributed work [42]. Reputation may load only on events carrying nonrecoverable exposure: capital locked for meaningful duration, fees paid to external parties, randomly selected audits survived, challenges survived where the challenger posted a bond burned on frivolous initiation—so a staged self-challenge costs the coalition the burn, by the logic of Corollary 3—losses absorbed, disputes adjudicated by parties whose fees are nonrecoverable to the coalition (the premise already priced in Theorem 1(a)), and time-weighted exposure without failure. "Independent" initiation and adjudication are themselves endogenously uncertifiable (Theorem 1), so challenge and dispute events carry reputational weight only through their nonrecoverable component—burned challenger bonds, external adjudicator fees—never through a certified independence claim. Two outputs must remain distinct: *Sybil-resistance weight* (how much irrecoverable exposure backs an identity—purchasable by burning money) and a *quality estimate* (how predictive the history is—which purchased exposure must not automatically improve). E itself becomes an *estimated* quantity—funding-source independence, requester diversity, cross-venue continuity, counterparty concentration, cyclic payment patterns—with collateral sized against the estimator's residual error. Such a layer estimates coalition risk, requires capital against the uncertainty, and settles claims when the estimate fails; it does not prove that counterparties are independent.

9 Related work

Douceur [2] established that free pseudonyms make endogenous identity insufficient; our Theorem 1 is the demand-side instance, and Remark 10 inherits his boundary. Mechanism design met the same obstacle in auctions: Yokoo, Sakurai, and Matsubara [40] showed that false-name bids break the incentive properties of combinatorial auctions and that no false-name-proof combinatorial auction protocol is Pareto efficient. Self-dealt demand is a false-name bid on the demand side of a work market. Cheng and Friedman [41] proved that no symmetric reputation function is Sybil-proof while asymmetric functions anchored at trusted nodes can be, the reputation-side counterpart of the trust-root escape in Remark 10. Seuken and Parkes [42] studied Sybil-proof accounting of contributed work in decentralized work-sharing networks; their accounting problem is close to ours, with service priority in place of consensus weight. Budish [3] bounds consensus security by the net economic cost of attack from the resource-flow side, with honest participation pinned down by a free-entry condition; Lemma 1 supplies the demand-side dual—which nominal flows must be *excluded* from that cost—and Proposition 2 extends his equilibrium logic to salvage-bearing work, where the same free-entry condition determines whether the attacker's salvage discount is offset by expanded honest supply. Bonneau [4] showed attack cost falls below assumed levels when consensus resources can be *rented*; we show it falls when demand payments can be *recovered*—the same shape applied to a different object, and the two compose. Carlsten et al. [5] exhibit instability when election rests on fee flows alone, adjacent to our cheap-weight floor. Dotan and Tochner [6] proved the computational obstructions to useful mining (private information, easy instances, external markets); our salvage deduction (2) is the economic completion of their result, and we state it as such rather than attributing the cost bound to them. Long-range attacks on proof of stake [8] are the historical-simulation analogue of Theorem 1, resolved by weak subjectivity—again imported scarcity. The verifier's dilemma [7] governs the audit economics of the settlement layer but is orthogonal to the impossibility. Empirical wash-trading detection [9, 10] instantiates the estimator of Theorem 1(b).

A recent economics-of-PoUW literature addresses the salvage axis directly, and we position against it explicitly. Bar-On, Komargodski, and Weinstein [25] (prior work) analyze equilibrium mining with external utilities (heterogeneous costs, exogenous external rewards); their verdict is *conditional*, and their own analysis identifies that concentrated access to those utilities can let a single actor control the network cheaply and that low access entropy reduces security—the centralization mode we describe in Remark 6, left there as an open threat with a liquidity-based mitigation we instead answer with the demand-independent floor. Pass [26] (concurrent)[1] develops a competitive-equilibrium model of PoUW with exogenous inference demand in which the economic cost of majority attack remains tied to the block reward across regimes—a genuine neutrality result. The gap our analysis fills against both is on the attack-cost side: their models assume exogenous demand and (for Pass) elastic supply, whereas Propositions 1–2 open the endogenous-demand (self-dealt), attacker-asymmetric-salvage, and inelastic-supply channels (Remark 4). A systematization of knowledge [27] reviews over fifty constructions with a formal economic model of miner incentives and external market value, concluding that economic and societal utility do not contribute to the security budget—convergent, from the survey side, with our impossibility-side account of *why* they cannot: recoverable transfers and unprovable demand exogeneity. Its informal threat analysis (§7.2) sketches an adversary who mines, colludes, and *trades in the external market* where task outputs carry value—including wash trading and cross-market leverage that funds an onchain attack from off-chain positions; these are the qualitative counterparts of the salvage

[1] The preprint discloses that the work was done while consulting for Pearl Research Labs.

term $V^{\mathcal{A}}_{\text{salvage}}$, the self-dealt demand of Lemma 1, and the funded attacker of Remark 6, which §3 prices. The transaction-fee-mechanism literature contains the direct analogue of Lemma 1: fake bids and self-dealt transactions are costless to their creator except for burned fees, and burning is what restores incentive properties [28, 29]; proof of burn [30] constructs a Sybil-resistant primitive from provably destroyed value. Corollary 3 specializes that established logic to demand-weighted leader election; the contribution is the explicit reduction and its scope conditions, not priority over fake-bid or proof-of-burn reasoning. The attestation-native design of [24] independently prices its bounded boost by fee burn at the mechanism level.

Early useful-work systems also narrow the architectural novelty. Primecoin made prime-chain search itself the consensus work [14]; Permacoin constructed a block-hash-bound storage puzzle so that the useful resource remained the proof-of-work resource [15]. Neither depends on customer-payment volume. Gridcoin later separated the two functions in deployed form: stake determines block-production odds, while BOINC work affects rewards only, and consensus is documented to survive the disappearance of BOINC demand [16]. Our contribution is not to originate work–election coupling or settlement–election separation, but to explain, using the demand-exogeneity and recoverability axes, why customer receipts cannot replace the independent floor.

Within PoUW, the property literature [12, 17, 18] named the verifiability–usefulness tension theoretically—Pearl [22] is its reported measured instance—and the systematic review [19] catalogues verification and hardness requirements; none formalizes self-dealt demand. Ofelimos [13] secures mining by stochastic local search with a miner-generated, hash-derived fallback when client-posted instances are absent—in our terms, a demand-independent work floor conceding Proposition 4's path by construction; REM [20] anchors in SGX—abandoning Assumption 2; recent inference-consensus proposals [23] and near-optimal certification of the matrix-multiplication primitive [21] strengthen Exec while leaving E untouched; proof-of-learning [43] attests machine-learning training as Exec evidence, and the spoofing attacks against it [44] show how hard Exec is for stochastic workloads before E arises at all; and an attestation-native stake-times-reputation design [24] independently converges on the floor-plus-priced-boost architecture (Remark 16). To our knowledge, the combination that survives contact with this literature is: the recoverable-transfer accounting of demand (Lemma 1), endogenous-demand indistinguishability (Theorem 1), the zero-demand-path floor for demand-derived weight (Proposition 4) with floor-relative sizing of free and burn-priced boosts (Corollaries 4 and 5), and the payment-tax reduction of demand-funded security to explicit burn (Corollary 3); the salvage-equilibrium analysis of §3 is our static, frozen-supply complement to [25, 26]—an explicit upper bound on the honest-exit equilibrium matching cost—with the delta stated in Remark 4. We claim the intersection and the named axes, not the equilibrium program.

10 Conclusion

The useful-work literature has been decomposing, slowly and expensively, because it conflated four goods bought with one security budget: consensus security, execution integrity, demanded-job binding, and output acceptance. Work certification is now near-optimal for the matrix-multiplication primitive under stated assumptions; job binding is cryptographically achievable; acceptance is a contractibility question. The residue is demand exogeneity—not derivable from endogenous receipts under free pseudonym creation and the stated observation model, however good the proofs of execution become.

The results have different scopes. A recoverable payment adds nothing to the attacker's cost, so no attack-cost lower bound may count it; genuine payments may still fund honest participation

while demand persists (Lemma 1, Proposition 2). Useful output discounts the attacker's unit cost unconditionally (Proposition 1). Within the stylized free-entry model, that discount degrades *security* under sufficient conditions—attacker-asymmetric salvage below the demand threshold, or inelastic honest supply (Proposition 2)—while whether deployed systems occupy those regimes remains an empirical conjecture (Remark 4). Receipts cannot prove that the payer is economically independent (Theorem 1). Even genuine demand is contingent, and P4–P5 require the mechanism to survive a projectable zero-demand path; the consensus floor therefore cannot rest on the demand component of receipts (Proposition 4). Demand intended to fund security above the floor should do so through an explicit burned tax instead of nominal volume (Corollary 3), with the resulting boost capped against the floor alone, because a coalition can still pay the burn after honest demand is gone (Corollary 5). A receipt can prove work and delivery, but not economically independent demand. A bounded receipt component may add security through an explicit burn or a separately proved non-scalar resource, while the leader-election floor rests on costs that no coalition can pay to itself.

References

[1] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008.

[2] J. R. Douceur. The Sybil attack. In *IPTPS*, 2002.

[3] E. Budish. Trust at scale: The economic limits of cryptocurrencies and blockchains. *Quarterly Journal of Economics*, 140(1):1–62, 2025. DOI: 10.1093/qje/qjae033. Earlier version: The economic limits of Bitcoin and the blockchain, NBER Working Paper 24717, 2018.

[4] J. Bonneau. Why buy when you can rent? Bribery attacks on Bitcoin-style consensus. In *Financial Cryptography Workshops (BITCOIN)*, 2016.

[5] M. Carlsten, H. Kalodner, S. M. Weinberg, and A. Narayanan. On the instability of Bitcoin without the block reward. In *ACM CCS*, 2016.

[6] M. Dotan and S. Tochner. Proofs of useless work: Positive and negative results for wasteless mining systems. arXiv:2007.01046, 2020.

[7] L. Luu, J. Teutsch, R. Kulkarni, and P. Saxena. Demystifying incentives in the consensus computer. In *ACM CCS*, 2015.

[8] V. Buterin. Proof of stake: How I learned to love weak subjectivity, 2014.

[9] L. W. Cong, X. Li, K. Tang, and Y. Yang. Crypto wash trading. *Management Science*, 2023.

[10] F. Victor and A. M. Weintraud. Detecting and quantifying wash trading on decentralized cryptocurrency exchanges. In *The Web Conference (WWW)*, 2021.

[11] O. Hart and J. Moore. Incomplete contracts and renegotiation. *Econometrica*, 56(4), 1988.

[12] M. Ball, A. Rosen, M. Sabin, and P. N. Vasudevan. Proofs of useful work. ECCC TR17-203, 2017.

[13] M. Fitzi, A. Kiayias, G. Panagiotakos, and A. Russell. Ofelimos: Combinatorial optimization via proof-of-useful-work. In *CRYPTO*, 2022.

[14] S. King. Primecoin: Cryptocurrency with prime number proof-of-work. Technical paper, `https://primecoin.io/primecoin-paper.pdf`, 2013.

[15] A. Miller, A. Juels, E. Shi, B. Parno, and J. Katz. Permacoin: Repurposing Bitcoin work for data preservation. In *IEEE Symposium on Security and Privacy*, pages 475–490, 2014. DOI: 10.1109/SP.2014.37.

[16] Gridcoin Project. Staking and Gridcoin FAQ. Project documentation, `https://gridcoin.us/wiki/staking.html` and `https://gridcoin.us/wiki/faq.html`, accessed 2026.

[17] F. Hoffmann. Challenges of proof-of-useful-work (PoUW). In *IEEE iGETblockchain*, 2022.

[18] A. Merlina, T. Garrett, and R. Vitenberg. On replacing cryptopuzzles with useful computation in blockchain proof-of-work protocols. *ACM Computing Surveys*, 2026. DOI: 10.1145/3828717. Preprint: arXiv:2404.15735 (2024).

[19] F. Bakhshi, M. Ashtiani, and M. Abdollahi Azgomi. A systematic literature review on proof-of-useful work consensus algorithms. *Blockchain: Research and Applications*, 2025. DOI: 10.1016/j.bcra.2025.100387.

[20] F. Zhang, I. Eyal, R. Escriva, A. Juels, and R. van Renesse. REM: Resource-efficient mining for blockchains. In *USENIX Security*, 2017.

[21] I. Komargodski, I. Schen, and O. Weinstein. Proofs of useful work from arbitrary matrix multiplication. arXiv:2504.09971; IACR ePrint 2025/685, 2025.

[22] A. Basu. The usefulness gap in proof-of-useful-work: An empirical study of Pearl's cuPOW protocol. arXiv:2606.04819 (v2), 2026. Preprint.

[23] L. Jimenez, M. Weatherspoon, B. Shen, Y. Sheng, J. Liu, and B. Li. HadAgent: Harness-aware decentralized agentic AI serving with proof-of-inference blockchain consensus. arXiv:2604.18614, 2026.

[24] S. Stefanović. Proof of useful attestation: A consensus primitive for attestation-native chains. arXiv:2605.25844, 2026.

[25] Y. Bar-On, I. Komargodski, and O. Weinstein. Proof of work with external utilities. arXiv:2505.21685, 2025.

[26] R. Pass. The economics of proof-of-useful-work. arXiv:2606.06700, 2026.

[27] P. Dikshit, A. Emami, J. Sedlmeir, and G. Fridgen. SoK: Is proof-of-useful-work really useful? IACR ePrint 2025/1814, 2025.

[28] T. Roughgarden. Transaction fee mechanism design. In *ACM Conference on Economics and Computation (EC)*, 2021.

[29] H. Chung and E. Shi. Foundations of transaction fee mechanism design. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2023.

[30] K. Karantias, A. Kiayias, and D. Zindros. Proof-of-burn. In *Financial Cryptography and Data Security (FC)*, 2020.

[31] Protocol Labs. Filecoin specification: storage power consensus and miner collaterals. https://spec.filecoin.io/, 2022.

[32] Filecoin Project. Verified deals and the Filecoin Plus (Fil+) 10× quality multiplier. Filecoin documentation, https://docs.filecoin.io/storage-providers/filecoin-deals/verified-deals, accessed 2026.

[33] Filecoin Project. FIP-0003: Filecoin Plus (notaries, root key holders, DataCap). https://github.com/filecoin-project/FIPs/blob/master/FIPS/fip-0003.md, 2020.

[34] Filecoin notary governance. Issue #811: recommendation for removal of notaries for abusing Filecoin Plus. https://github.com/filecoin-project/notary-governance/issues/811, 2023.

[35] Bittensor. How Yuma Consensus 3 makes Bittensor more fair; and analysis of the weight-copying failure mode. Bittensor documentation, https://docs.learnbittensor.org/yc3-blog and https://docs.learnbittensor.org/papers/BT_Weight_Copier-29May2024.pdf, accessed 2026.

[36] Helium Systems, Inc. Helium: a decentralized wireless network. Whitepaper, http://whitepaper.helium.com/, 2018.

[37] Helium Community. HIP-70: scaling the Helium network (migration to Solana). Helium Improvement Proposals, https://github.com/helium/HIP/blob/main/0070-scaling-helium.md, 2022.

[38] LayerZero Labs. Addressing sybil activity. LayerZero Official (Medium), https://medium.com/layerzero-official/addressing-sybil-activity-a2f92218ddd3, May 2024. Description of the self-report and community-bounty process.

[39] Hop DAO. HIP-30: HOP airdrop sybil-hunter distribution. Hop Protocol governance forum, https://forum.hop.exchange/t/rfc-hop-airdrop-sybil-hunter-distribution/899, 2022–2023.

[40] M. Yokoo, Y. Sakurai, and S. Matsubara. The effect of false-name bids in combinatorial auctions: New fraud in Internet auctions. *Games and Economic Behavior*, 46(1):174–188, 2004. DOI: 10.1016/S0899-8256(03)00045-9.

[41] A. Cheng and E. Friedman. Sybilproof reputation mechanisms. In *Third Workshop on Economics of Peer-to-Peer Systems (P2PEcon '05)*, co-located with ACM SIGCOMM, 2005.

[42] S. Seuken and D. C. Parkes. Sybil-proof accounting mechanisms with transitive trust. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 205–212, 2014.

[43] H. Jia, M. Yaghini, C. A. Choquette-Choo, N. Dullerud, A. Thudi, V. Chandrasekaran, and N. Papernot. Proof-of-learning: Definitions and practice. In *IEEE Symposium on Security and Privacy*, 2021.

[44] R. Zhang, J. Liu, Y. Ding, Z. Wang, Q. Wu, and K. Ren. "Adversarial examples" for proof-of-learning. In *IEEE Symposium on Security and Privacy*, 2022.

[45] X. Chen, C. Papadimitriou, and T. Roughgarden. An axiomatic approach to block rewards. In *ACM Conference on Advances in Financial Technologies (AFT)*, 2019.