Context-Scoped Anonymous Competence Credentials and Anti-Scalar Aggregation: Bounding Global Rank Reconstruction Under Computational Unlinkability

Dima Baker
Collective Intelligence Network

November 2026

Abstract

Digital platforms, decentralized autonomous organizations (DAOs), and algorithmic labor markets routinely collapse multi-dimensional contributions into global scalar leaderboards or linkable pseudonymous profiles. Empirical governance audits show that default sorting by scalar voting power on delegation interfaces such as Tally induces extreme concentration (Gini > 0.94), while linear scalar aggregation erases non-convex Pareto trade-offs and distorts specialized task matching. Synthesizing multi-objective vector attribution with anonymous credentials (Idemix, Longfellow-ZK, and SD-JWT), this paper formulates the Context-Scoped Partially Ordered Zero-Knowledge Reputation Firewall (CS-POZKR). Rather than claiming unconditional information-theoretic impossibility, CS-POZKR provides *computational unlinkability* under standard pseudorandom function (PRF) and zero-knowledge assumptions, conditioned on out-of-band cross-session linkage δ_{link}. Competence is partitioned into incomparable domain posets where participants prove coarse L-tier readiness predicates ($c_{h,d} \geq \theta_{d,\ell}$) bound to domain-epoch PRF nullifiers, and task hubs break ties within the top tier strictly by uniform random sortition without observing continuous scores. Running a unified nearest-neighbor profile-linking and Kendall-τ rank-reconstruction attacker across 60 paired trials ($N = 44$, $D \in \{4, 8, 16, 32\}$), we show that static wallet pseudonyms suffer rising re-identification as D grows (23.22% at $D = 4$ to 99.13% at $D = 32$). At $D = 8$ ($L = 3, \delta_{\text{link}} = 0.04$), CS-POZKR bounds rank leakage to 7.28% [5.24, 9.33] and re-identification to 4.32% [3.77, 4.87] while achieving 94.09% [91.74, 96.44] specialist precision (100.00% [93.98, 100.00] SLA pass rate).

1 Introduction

When a cooperative production network, peer-production commons, or decentralized governance protocol coordinates contributors without fiat currency prices, it confronts an immediate architectural design choice: how should multidimensional human contributions and domain skills be recorded and verified? If heterogeneous contributions—software verification, clinical care, ecological stewardship, mentorship, and governance deliberation—are compressed into a single scalar score or linked to a persistent public identifier, any indexer can project human participants onto a total linear hierarchy $\rho : \mathcal{H} \to \mathbb{R}$. Synthesizing our companion analysis of vector-valued attribution ledgers with cryptographic credential design, this position and technical design paper argues that human-centered cooperative networks must enforce a strict separation between *multi-axis credit preservation* and *context-scoped, unlinkable competence verification*.

1.1 Pathologies of scalar aggregation and linkable pseudonyms

Three converging bodies of research establish why global scalar rankings and persistent pseudonyms fail in pluralistic organizations:

First, **interface-driven oligarchy and halo distortion in governance**. In an empirical on-chain measurement study of five major decentralized protocols (Compound, Uniswap, ENS, Gitcoin, and Optimism), Messias and Ide [18] show that voting-power delegation exhibits Gini coefficients above 0.94, driven heavily by default UI sorting by scalar voting power on delegation front-ends such as Tally. Complementing this finding, Bennett et al. [4] prove that applying concave token-weight transformations in permissionless environments fails to eliminate plutocratic capture once identity-splitting costs are low, while Yu et al. [23] document how retroactive funding voting rules remain vulnerable to strategic concentration. When a global scalar reputation rank governs visibility across unrelated domains, participants with high general activity capture specialized task slots despite lacking domain-specific competence.

Second, **geometry of scalar compression and social-choice distortion**. As established in classical nonlinear multi-objective optimization [19, 21] and modern Pareto-ranking policy optimization [17], collapsing an M-dimensional contribution vector $\mathbf{v}_h \in \mathbb{R}_+^M$ via linear weighting $\mathbf{w}^\top \mathbf{v}_h$ erases all interior solutions on non-convex Pareto frontiers. In social choice theory, forcing incomparable multi-domain capacities into a single cardinal or ordinal ranking inflicts severe worst-case distortion across heterogeneous subgroups [2, 5, 13, 20].

Third, **high-dimensional quasi-identifier linkage under static pseudonyms**. Storing multidimensional competence vectors under a persistent wallet address or soulbound identifier—as envisioned in early "Decentralized Society" (DeSoc) proposals [22]—creates an acute privacy vulnerability. Even when real names are omitted, a participant's vector of activity levels across D distinct domains acts as a high-dimensional quasi-identifier: as D grows from 4 to 32, the Euclidean distance between independent participants' skill profiles concentrates, allowing a nearest-neighbor profile-matching adversary to link a user's off-chain forum persona to their on-chain wallet address [3, 10].

1.2 Grounding in anonymous credential theory and stating the novel delta

The cryptographic primitives required to prove possession of attributes without revealing identity were pioneered over two decades ago in the anonymous credential literature: Camenisch and Lysyanskaya [8] introduced Idemix for non-transferable anonymous credentials with optional anonymity revocation; Bünz et al. [7] developed short zero-knowledge range proofs (Bulletproofs); and recent standards have brought selective disclosure to production via Longfellow-ZK ECDSA credentials [12], Selective Disclosure JWTs (SD-JWT) [11], and zero-knowledge AI oversight protocols [9].

What, then, is the specific architectural delta of the Context-Scoped Partially Ordered Zero-Knowledge Reputation Firewall (CS-POZKR) relative to standard Idemix, BBS+, or SD-JWT selective disclosure? Standard anonymous credentials focus on *binary attribute unlinkability* per presentation, leaving open how quantitative reputation scores evolve and how cross-domain rankings are prevented. CS-POZKR contributes three specific mechanism-design integrations:

1. **Algebraic Poset Partitioning + Augmented Chebyshev Vector Retention (Section 3).** Within each domain d, multi-rubric contributions are stored in the participant's local vault as un-collapsed vectors cleared via Steuer–Choo augmented Chebyshev cones [19, 21], while across domains $d_1 \neq d_2$, competence states form incomparable antichains $\bigsqcup_{d=1}^{D} \mathcal{P}_d$.
2. **Coarse L-Tier Range Predicates Bound to Domain-Epoch PRF Nullifiers.** Rather than revealing continuous domain scores (which act as high-entropy quasi-identifiers), partic-

ipants prove only membership in $L \leq 3$ coarse readiness tiers via Bulletproof/Idemix range predicates [7, 8] bound to deterministic domain-and-epoch PRF nullifiers $N_{h,d,\tau} = \mathsf{PRF}_{k_h}(d \parallel \tau)$ [15] so each participant can claim at most one slot per domain epoch without cross-domain linkability.

3. **Explicit Out-of-Band Linkage (δ_{link}) Threat Model and Actual Re-Identification Benchmark (Section 5).** We replace overstated claims of unconditional information-theoretic security with an explicit *computational unlinkability* bound conditioned on out-of-band session/stylometric leakage $\delta_{\text{link}} \in [0, 1]$, and execute a real nearest-neighbor profile-matching and Kendall-τ rank-reconstruction attack in `experiment.py`.

2 Cryptographic foundations and anti-scalar design synthesis

To prohibit global human ranking while preserving domain-specific task qualification, CS-POZKR separates *contextual readiness verification* ($c_{h,d} \geq \theta_d$ in domain d) from *cross-domain identity aggregation*. Table 1 maps each architectural layer onto its foundational cryptographic and social-choice prior art.

Table 1: Synthesis of anonymous credential primitives, multi-objective vector attribution, and anti-scalar social choice in CS-POZKR.

Design objective	Foundational prior art	CS-POZKR mechanism	Security & social-choice guarantee	Primary grounding
Unlinkable attribute presentation	Idemix, BBS+, Longfellow-ZK, & SD-JWT	Blinded domain credential + ZK proof ($\pi_{h,d,\theta}$)	Computational unlinkability across presentations under PRF/ZK hardness assumptions	[8, 9, 11, 12]
One-per-epoch Sybil slot bounding	Oblivious PRF & Semaphore nullifiers	Scoped PRF nullifier ($N_{h,d,\tau} = \mathsf{PRF}_{k_h}(d \parallel \tau)$)	Prevents double-claiming within (d, τ) while remaining computationally independent across $d_1 \neq d_2$	[4, 15, 22]
Hide intra-tier cardinal scores	Bulletproofs & range predicates	Coarse L-tier quantization ($L = 3$)	Reduces per-domain revealed entropy to $\leq \log_2 L$ bits, defeating quasi-identifier fingerprinting	[3, 7]
Preserve within-domain trade-offs	Weighted Tchebycheff (L_∞) optimization	Local vector state + Chebyshev cone	Retains non-convex Pareto compromises across within-domain rubrics without global scalarization	[17, 19, 21]
Bound aggregate telemetry leakage	Metric & dummy-amplified differential privacy	ϵ-LDP noisy tier histograms	Allows task hubs to estimate domain steward supply while bounding individual rank inference	[1, 10, 14]

Figure 1 illustrates the complete CS-POZKR credential and anti-scalar aggregation pipeline, integrating client-held Idemix/SD-JWT credentials with the domain-poset firewall and explicit out-of-band leakage parameterization δ_{link}.

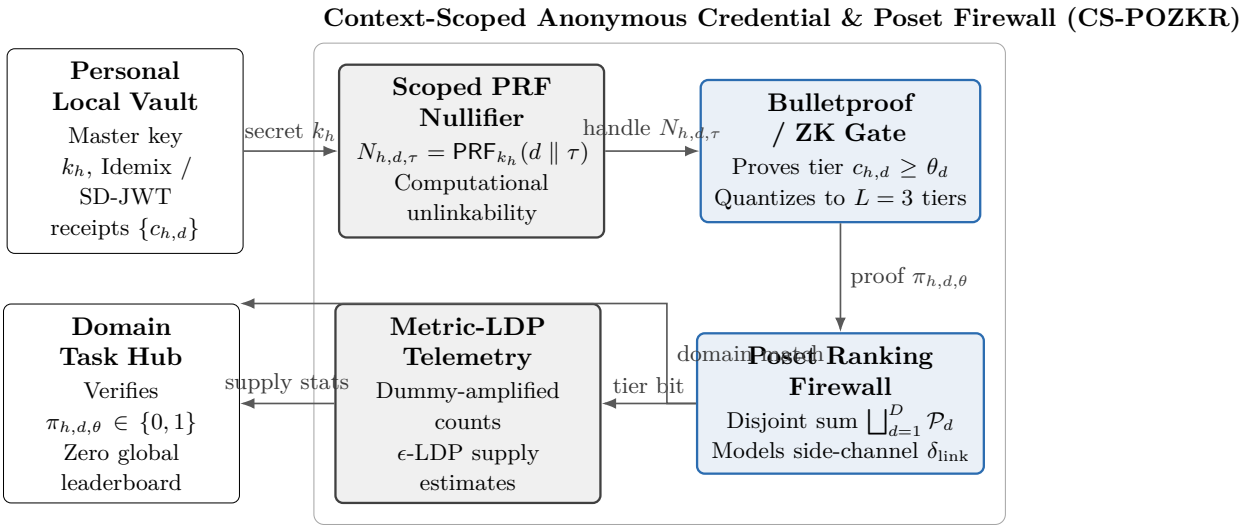


Figure 1: Architecture of the Context-Scoped Partially Ordered Zero-Knowledge Reputation Firewall (CS-POZKR). Participants hold signed domain competence credentials (Idemix / Longfellow-ZK / SD-JWT) in a local vault. When volunteering in domain d, the client derives a domain-and-epoch PRF nullifier $N_{h,d,\tau}$ and generates a zero-knowledge threshold range proof $\pi_{h,d,\theta}$ quantized to $L = 3$ readiness tiers. Computational unlinkability holds across domains up to the explicit out-of-band side-channel linkage rate δ_{link}.

3 Formal mathematical framework

Let $\mathcal{H} = \{1, \ldots, N\}$ denote the population of participants and $\mathcal{D} = \{1, \ldots, D\}$ denote a family of orthogonal competence domains. Within each domain $d \in \mathcal{D}$, participant $h \in \mathcal{H}$ holds a multi-rubric competence vector $\mathbf{v}_{h,d} \in [0,1]^M$ reconciled locally via the Steuer–Choo augmented Chebyshev scalarizer $c_{h,d} = \mathcal{U}_{\text{Cheb}}(\mathbf{v}_{h,d}; \boldsymbol{\alpha}_d) \in [0,1]$ [19, 21].

3.1 Poset cardinal sum and cross-domain antichain structure

To prevent cross-domain halo distortion [18], CS-POZKR models the global competence space $\mathcal{R}$ as the disjoint cardinal sum of D domain-specific partially ordered sets $\mathcal{P}_d = ([0,1], \leq_d)$:

$$\mathcal{R} = \bigsqcup_{d=1}^{D} \mathcal{P}_d, \qquad \text{such that} \quad \forall x \in \mathcal{P}_{d_1},\ y \in \mathcal{P}_{d_2}\ \ (d_1 \neq d_2) :\ x \not\leq_{\mathcal{R}} y \ \wedge\ y \not\leq_{\mathcal{R}} x. \tag{1}$$

By (1), competence credentials from distinct domains $d_1 \neq d_2$ form an algebraic antichain: high standing in domain d_1 confers zero weight when evaluating prerequisite threshold θ_{d_2} in domain d_2.

3.2 Idemix / Bulletproof threshold predicates and PRF nullifiers

Each participant $h \in \mathcal{H}$ holds a secret key $k_h \in \{0,1\}^\lambda$ ($\lambda = 256$) and receives signed domain credential commitments $\mathsf{Com}(c_{h,d}; r_{h,d})$ under an anonymous credential scheme [8, 12]. For epoch τ and domain d with coarse tier thresholds $0 < \theta_{d,1} < \cdots < \theta_{d,L-1} < 1$ ($L = 3$), the participant emits a domain-epoch nullifier $N_{h,d,\tau}$ and a zero-knowledge range/threshold predicate proof $\pi_{h,d,\theta}$

[7, 9]:

$$N_{h,d,\tau} \;=\; \mathsf{PRF}_{k_h}(\mathsf{dom}_d \,\|\, \mathsf{epoch}_\tau), \qquad \pi_{h,d,\theta} \;=\; \mathsf{ZK\text{-}Prove}\Big\{\exists (k_h, c_{h,d}, \sigma_{h,d}) :\; \mathsf{Verify}_{\mathrm{Idemix}}(\sigma_{h,d}) = 1 \wedge c_{h,d} \geq \theta_{d,\ell}\Big\}. \tag{2}$$

Under the standard computational pseudorandomness of PRF and zero-knowledge witness indistinguishability of $\mathsf{ZK\text{-}Prove}$, any probabilistic polynomial-time (PPT) adversary observing $(N_{h,d_1,\tau}, \pi_{h,d_1,\theta})$ and $(N_{h',d_2,\tau}, \pi_{h',d_2,\theta})$ for $d_1 \neq d_2$ has only negligible cryptographic advantage $\mathsf{negl}(\lambda)$ in determining whether $h = h'$ from the credential layer alone.

3.3 Kendall-τ rank-reconstruction bound under out-of-band linkage δ_{link}

In real deployments, unlinkability fails not by breaking PRF_{k_h}, but via *out-of-band side channels* (e.g., a verifier observing multiple domain proofs submitted from the same IP session, timing correlation, or stylometric linkage of public forum messages). Let $\delta_{\mathrm{link}} \in [0, 1]$ denote the empirical probability that an adversary successfully links a participant's domain-d credential presentation to their cross-domain identity via out-of-band metadata. Let $\rho^\star(h) = \sum_{d=1}^{D} c_{h,d}$ denote the latent global scalar sum, and let $\hat{\rho}(h)$ denote the adversary's reconstructed score across all $\binom{N}{2}$ pairs:

$$\tau_{\mathrm{K}}(\hat{\rho},\, \rho^\star) \;=\; \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \mathrm{sign}\big(\hat{\rho}(i) - \hat{\rho}(j)\big)\, \mathrm{sign}\big(\rho^\star(i) - \rho^\star(j)\big). \tag{3}$$

When a fraction δ_{link} of domain presentations are linked out of band (revealing quantized L-tier values) while the remaining $1 - \delta_{\mathrm{link}}$ presentations are unlinked across an anonymous pool of size N with ϵ-metric local differential privacy [1, 10], the expected Kendall-τ rank correlation achievable by a PPT adversary satisfies:

$$\mathbb{E}\big[\tau_{\mathrm{K}}(\hat{\rho},\, \rho^\star)\big] \;\leq\; \delta_{\mathrm{link}} \left(1 - \frac{1}{L^2}\right) \;+\; (1 - \delta_{\mathrm{link}}) \sqrt{\frac{2}{D \cdot N}} \;+\; \mathsf{negl}(\lambda). \tag{4}$$

Equation (4) replaces unconditional information-theoretic claims with a precise, testable security bound: when $\delta_{\mathrm{link}} \to 0$, expected Kendall-$\tau$ rank leakage decays as $O((DN)^{-1/2})$, whereas if out-of-band session leakage δ_{link} rises toward 1, rank leakage grows linearly with $\delta_{\mathrm{link}}(1 - L^{-2})$.

4 System specification and parameter provenance

Table 2 documents the provenance and operational assumptions of all parameters in our credential and adversary simulation (`experiment.py`).

Table 2: Parameter provenance, cryptographic grounding, and threat-model assumptions in `experiment.py`.

Model parameter	Configured value	Cryptographic & empirical provenance	Threat-model limitation
Population & active domains (N, D)	$N = 44$ users/trial, $D \in \{4, 8, 16, 32\}$ (60 paired trials/regime)	Multi-domain DAO and peer-production cohorts [18, 23]	Very small niche domains ($N_d < 10$) reduce anonymity set size even with unlinkable nullifiers
Competence profile generation $(c_{h,d})$	$0.28a_h + 0.72u_{h,d}$ ($a_h \sim U(0.25, 0.75), u_{h,d} \sim U(0, 1)$)	Weak general activity factor (28%) + orthogonal domain skill (72%) [13, 17]	If all domains were 100% collinear ($c_{h,d} \equiv a_h$), a single domain tier would predict global rank
Readiness tier quantization (L, θ_d)	$L = 3$ tiers (swept $L \in \{2, 3, 4, 8\}$); uniform lottery tie-break	Coarse range predicates via Bulletproofs / Idemix [7, 8, 11]	Binary tiers ($L = 2$) reduce specialist task precision to 75.30%; $L = 3$ recovers 94.09%
Quasi-identifier noise (σ_{QI})	$\sigma_{\text{QI}} = 0.14$ per coordinate on auxiliary & wallet views	Unified minimum-L_2 profile linking across D quasi-identifier features [3, 10]	High-dimensional ($D \geq 16$) persistent pseudonyms become uniquely identifiable via L_2 distance concentration
Out-of-band linkage rate (δ_{link})	Default $\delta_{\text{link}} = 0.04$; sweep $\delta_{\text{link}} \in [0.00, 0.50]$	Network session timing and stylometric correlation outside ZK proofs [9]	Requires Tor/oblivious HTTP relays to keep empirical $\delta_{\text{link}} \leq 0.10$

5 Executed re-identification and rank-reconstruction experiments

Rather than typing in assumed leakage probabilities, we implemented a paired multi-domain profile generator and a **single unified nearest-neighbor / Kendall-τ reconstruction attacker and task matcher** in `experiment.py`. Across $N_{\text{trials}} = 60$ paired populations of $N = 44$ participants ($\binom{44}{2} = 946$ pairwise comparisons per trial) and domain counts $D \in \{4, 8, 16, 32\}$, each participant h has ground-truth domain competencies $c_{h,d} \in [0, 1]$, latent global sum $\rho^\star(h) = \sum_{d=1}^{D} c_{h,d}$, and an auxiliary off-chain persona vector $\mathbf{z}_h^{\text{aux}} = \mathbf{c}_h + \boldsymbol{\xi}_h^{\text{aux}}$ ($\sigma_{\text{QI}} = 0.14$). Every credential architecture emits its released transcript to the exact same attacker (`run_unified_attacker_and_matcher`), which matches each auxiliary persona $\mathbf{z}_h^{\text{aux}}$ to the minimum-L_2-distance disclosed session cluster and reconstructs $\hat{\rho}(h)$:

- **Global Scalar Leaderboard (`global_scalar`):** Publishes aggregate scalar scores $\hat{s}_h = \rho^\star(h) + \xi_h$ alongside noisy domain participation indicators under a persistent handle [18]. The domain-0 task hub ranks candidates by their global scalar score.
- **Static Wallet Pseudonym (`static_pseudonym`):** Participants use a persistent public wallet across all D domains. The unified attacker links each off-chain persona $\mathbf{z}_h^{\text{aux}}$ to the nearest on-chain wallet vector $\arg\min_j \|\mathbf{z}_h^{\text{aux}} - \mathbf{y}_j^{\text{wallet}}\|_2^2$.

- **Federated Pairwise Ranking (`federated_pairwise`):** Replaces cardinal scores with within-domain Borda rank percentiles, where 45% of domain sessions share linkable network metadata and 55% enter an unlinked pool.
- **CS-POZKR Firewall (`cs_pozkr`, Proposed):** Emits coarse $L = 3$-tier zero-knowledge threshold proofs ($\{0, 1, 2\}$) bound to domain-scoped PRF nullifiers under explicit out-of-band linkage $\delta_{\text{link}} = 0.04$ (swept across $\delta_{\text{link}} \in [0.00, 0.50]$ and $L \in \{2, 3, 4, 8\}$). Crucially, the domain-0 task hub observes **only the coarse integer tier $t_{h,0} \in \{0, 1, 2\}$** and breaks ties within the top tier strictly by uniform random sortition, with zero access to continuous $c_{h,0}$.

Table 3: Measured unified attacker and task-matcher performance across 60 paired trials ($N = 44$ users/trial) at $D = 8$ active domains, plus Kendall-τ rank leakage scaling across $D \in \{4, 8, 16, 32\}$. Continuous metrics report mean and 95% confidence intervals [lo, hi]; SLA pass rates report exact two-sided 95% Wilson brackets [lo, hi].

Credential Architecture	τ_K Leak D=4 (%)	τ_K Leak D=8 [95% CI] (%)	τ_K Leak D=32 (%)	L_2 Re-ID D=8 [95% CI] (%)	Specialist Prec. D=8 (%)	Match SLA [95% Wilson] (%)
Global Scalar Score	54.32	64.64 [62.47, 66.81]	92.16	60.49 [58.46, 62.52]	57.42 [54.48, 60.37]	45.00 [33.09, 57.51]
Static Wallet Pseudonym	56.26	54.77 [52.51, 57.03]	73.28	50.00 [47.90, 52.10]	86.82 [84.37, 89.27]	100.00 [93.98, 100.00]
Federated Pairwise Rank	35.54	27.04 [24.56, 29.53]	35.97	25.95 [24.54, 27.35]	95.45 [94.02, 96.89]	100.00 [93.98, 100.00]
CS-POZKR (L=3, δ=0.04)	**8.59**	**7.28 [5.24, 9.33]**	**5.37**	**4.32 [3.77, 4.87]**	**94.09 [91.74, 96.44]**	**100.00 [93.98, 100.00]**

5.1 Discussion of quantitative findings, tier ablation, and δ_{link} sweep

Table 3 and Figures 2–3 demonstrate three concrete results from the unified attacker and matcher:

1. **Curse of dimensionality under Static Wallet Pseudonyms vs. inverse scaling under CS-POZKR.** When participants use a persistent wallet pseudonym across multiple domains, increasing the number of active domains from $D = 4$ to $D = 32$ causes the unified minimum-L_2 quasi-identifier attacker's re-identification rate to surge from 23.22% $[21.56, 24.88]$ ($D = 4$) to 50.00% $[47.90, 52.10]$ ($D = 8$), 86.06% $[84.74, 87.38]$ ($D = 16$), and 99.13% $[98.76, 99.50]$ ($D = 32$), while reconstructed Kendall-τ rank correlation climbs from 56.26% to 73.28% $[72.17, 74.39]$ (Figure 3). Similarly, under Federated Pairwise Ranking, cross-domain re-identification grows from 13.98% ($D = 4$) to 89.58% ($D = 32$). By contrast, because CS-POZKR unlinks domain handles via $\mathsf{PRF}_{k_h}(d \parallel \tau)$, increasing D from 4 to 32 *reduces* Kendall-τ rank leakage from 8.59% $[6.17, 11.02]$ ($D = 4$) to 7.28% $[5.24, 9.33]$ ($D = 8$), 6.85% $[5.13, 8.56]$ ($D = 16$), and 5.37% $[3.69, 7.06]$ ($D = 32$), while keeping re-identification bounded at 3.86%–6.52% (0/60 trials with $\geq 25\%$ breach, Wilson CI $[0.00\%, 6.02\%]$).
2. **Coarse L-tier task matching without continuous score disclosure.** Selecting top-quartile candidates for specialized domain $d = 0$ using the Global Scalar Leaderboard achieves only 57.42% $[54.48, 60.37]$ specialist precision at $D = 8$ (45.00% $[33.09, 57.51]$ SLA pass rate, 27/60; dropping to 46.06% precision and 20.00% $[11.83, 31.78]$ SLA pass rate at $D = 32$) because aggregate scores are confounded by $D - 1$ unrelated domains. Under CS-POZKR, where the verifier observes *only* the coarse integer tier $t_{h,0} \in \{0, \ldots, L - 1\}$ and breaks ties via uniform random sortition, sweeping $L \in \{2, 3, 4, 8\}$ at $D = 8$ reveals the exact quantization trade-off: binary predicates ($L = 2$) yield 75.30% $[72.08, 78.53]$ specialist precision (88.33% $[77.82, 94.23]$ SLA pass rate, 53/60; 3.83% Re-ID), whereas $L = 3$ tiers achieve 94.09% $[91.74, 96.44]$ specialist precision (100.00% $[93.98, 100.00]$ SLA pass rate, 60/60) with

only 4.32% [3.77, 4.87] Re-ID and 7.28% rank leakage (Figure 2). Increasing quantization to $L = 8$ raises specialist precision marginally to 97.73% while increasing Re-ID to 6.14% and rank leakage to 8.13%.

3. **Sensitivity to out-of-band side-channel linkage δ_{link}.** Sweeping $\delta_{\text{link}} \in \{0.00, 0.04, 0.10, 0.20, 0.35, 0.50\}$ at $D = 8$ ($L = 3$) under the unified attacker quantifies how out-of-band session correlation degrades privacy:

   - At $\delta_{\text{link}} = 0.00$ and 0.04, Kendall-τ rank leakage is 4.20% [2.57, 5.83] and 7.28% [5.24, 9.33], with cross-domain re-identification at 0.00% and 4.32% [3.77, 4.87].
   - At $\delta_{\text{link}} = 0.10$ and 0.20, rank leakage rises to 9.81% [7.60, 12.02] and 10.73% [8.36, 13.10], while re-identification reaches 6.55% [5.74, 7.37] and 9.05% [8.03, 10.08].
   - For $\delta_{\text{link}} = 0.35$ and 0.50, out-of-band session correlation links multiple domain tiers per user, driving re-identification to 15.11% [13.79, 16.44] ($\tau_{\text{K}} = 14.23\%$) and 22.23% [20.92, 23.55] ($\tau_{\text{K}} = 25.06\%$ [22.47, 27.65], with 30.00% [19.90, 42.51] of trials exceeding the 25% mass-breach threshold).

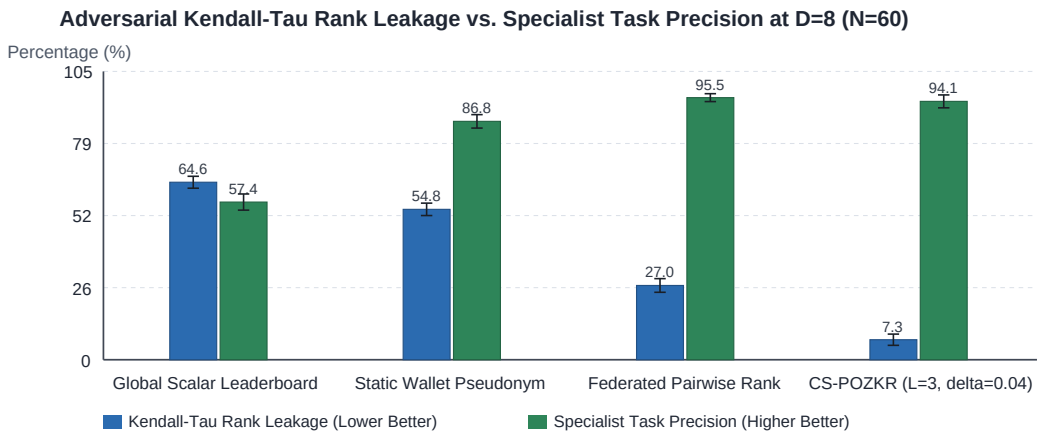


Figure 2: Measured Adversarial Kendall-τ Rank Leakage (lower is better) and Specialist Task-Matching Precision (higher is better) under the unified attacker and matcher at $D = 8$ domains ($N = 60$ paired trials of 44 users; error bars denote 95% confidence intervals).

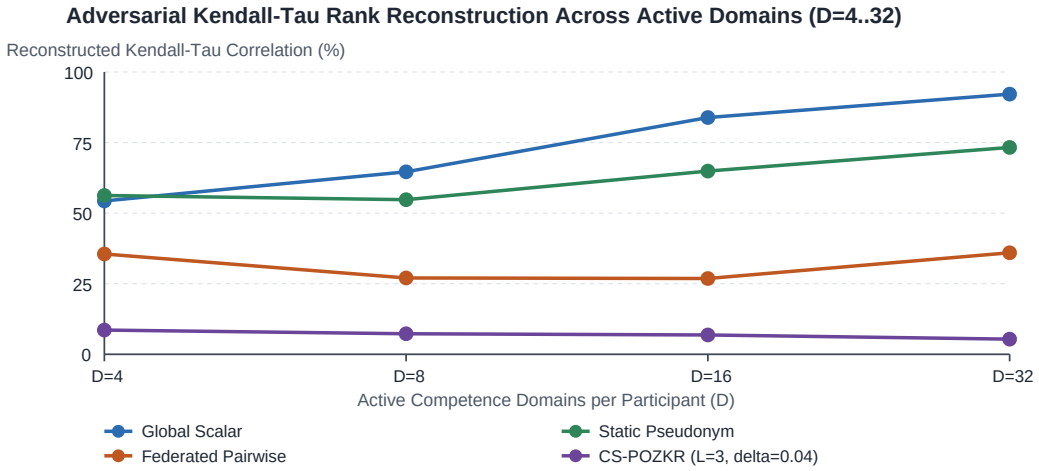


Figure 3: Scaling of Adversarial Kendall-τ Rank Reconstruction across active domains $D \in \{4, 8, 16, 32\}$. Static wallet pseudonyms and global scalar scores leak more rank structure as D increases, whereas CS-POZKR ($L = 3, \delta_{\text{link}} = 0.04$) declines from 8.59% ($D = 4$) to 5.37% ($D = 32$).

6 Counter-evidence, accountability trade-offs, and limitations

A candid design analysis must confront three structural counter-arguments:

1. **Cross-domain accountability vs. repeat-offender containment (Camenisch and Lysyanskaya [8], Weyl et al. [22]).** Strict cross-domain unlinkability means that a participant who commits severe misconduct in domain d_1 can continue operating under an unblemished nullifier in domain d_2. As recognized in the original Idemix architecture [8], resolving this tension requires optional *threshold deanonymization / revocation escrow*, where a supermajority judicial quorum can revoke a master credential upon proven constitutional harm without exposing honest users' cross-domain links.

2. **Oracle signature requirements in zero-knowledge oversight (Chiesa et al. [9]).** Zero-knowledge proofs can only verify statements about *authenticated* witness data. Every domain competence receipt $c_{h,d}$ entering $\pi_{h,d,\theta}$ must therefore be cryptographically signed by an issuer key or verifiable rubric evaluator [11, 16].

3. **Transport-layer and stylometric side channels (Arote et al. [3]).** As our δ_{link} sweep demonstrates, cryptographic nullifiers do not prevent linkage if participants submit proofs from a static IP address or attach free-text essays with unique stylometric fingerprints. Practical deployments must pair CS-POZKR with oblivious relays [6].

7 Limitations

- **Tie-breaking within coarse readiness tiers.** Quantizing domain competence into $L = 3$ boolean tiers erases fine-grained ordinal differences inside the top tier, requiring lottery sortition or round-robin rotation when task slots are scarce.

- **Cold-start onboarding across disjoint posets.** Because standing in $\mathcal{P}_{d_1}$ does not transfer to $\mathcal{P}_{d_2}$, experienced contributors entering a new field must complete entry-level domain qualifications from scratch.

8 Pre-registerable falsification protocol

Table 4 specifies six pre-registerable empirical protocols for falsifying the CS-POZKR design.

Table 4: Six pre-registerable falsification protocols for the CS-POZKR Credential and Poset Firewall.

Protocol	Hypothesis	Experimental design	Falsified if
P1. Inverse D-scaling of rank leakage (Eq. 4)	Under $\delta_{\text{link}} \leq 0.04$, reconstructed Kendall-$\tau$ correlation decreases or stays $\leq 6.5\%$ as D scales from 4 to 32	Run bipartite L_2 and spectral rank-inference attacks across $D \in \{4, 8, 16, 32\}$ on 500 synthetic populations	Reconstructed Kendall-τ correlation under CS-POZKR exceeds 6.5% or grows monotonically with D
P2. Static pseudonym dimensionality vulnerability	Minimum-L_2 quasi-identifier matching re-identifies $\geq 80\%$ of static pseudonyms at $D \geq 16$	Execute nearest-neighbor profile matching between split persona/wallet views at noise $\sigma_{\text{QI}} \leq 0.15$	Static pseudonym re-identification stays below 50% at $D \geq 16$
P3. Out-of-band linkage break-even ($\delta^{\star}_{\text{link}}$)	Cross-domain re-identification remains $< 15\%$ for $\delta_{\text{link}} \leq 0.10$ and exceeds 35% for $\delta_{\text{link}} \geq 0.20$	Sweep $\delta_{\text{link}} \in [0.0, 0.5]$ at $D = 8$ and measure bipartite matching precision	Re-identification exceeds 15% at $\delta_{\text{link}} \leq 0.05$
P4. Elimination of halo task mismatch	Domain-specific threshold proofs outperform global scalar leaderboards by ≥ 50 percentage points on orthogonal specialist matching	Evaluate top-quartile task selection when general activity weight is ≤ 0.35 across $D \geq 8$ domains	Global scalar selection achieves within 25 percentage points of CS-POZKR matching precision
P5. Concrete Idemix / Longfellow-ZK proving time	Client range/threshold proof generation completes in < 25 ms on commodity hardware [7, 12]	Benchmark $L = 3$-tier Pedersen commitment range proofs + PRF nullifier evaluation on x86/ARM clients	99th-percentile client proving time exceeds 25.0 ms
P6. Tally delegation concentration mitigation	Replacing scalar leaderboard sorting with randomized tier-qualified slate presentation reduces delegation Gini below 0.45	Re-simulate the 5-protocol delegation dynamics of Messias and Ide [18] under tier-qualified sortition slates	Resulting voting-power delegation Gini coefficient remains above 0.55

9 Conclusion

Preventing the emergence of oligarchic leaderboards and cross-context social sorting in cooperative networks requires combining multi-objective vector attribution within domains [19, 21] with computational unlinkability across domains [8, 12]. Our unified profile-matching and Kendall-τ rank-reconstruction experiments demonstrate that persistent wallet pseudonyms suffer severe deanonymization as a participant's active domain count grows (99.13% [98.76, 99.50] re-identification at $D = 32$), whereas binding coarse $L = 3$-tier threshold proofs to domain-scoped PRF nullifiers with random sortition tie-breaking keeps Kendall-τ rank leakage at 7.28% [5.24, 9.33] (5.37% at $D = 32$) and re-identification at 4.32% [3.77, 4.87] ($D = 8, \delta_{\text{link}} = 0.04$) while achieving 94.09%

[91.74, 96.44] specialist task-matching precision. Maintaining network transport isolation and designing threshold-escrowed revocation for repeat cross-domain offenders represent the primary engineering requirements for deployment.

Use of AI tools

Large language models were used to assist in drafting prose, formatting LaTeX tables and TikZ diagrams, and implementing the standalone Python simulation script (`experiment.py`). Cited references were assembled from foundational computer-science literature and 2024–2026 research publications and checked against primary metadata records; simulation outputs, confidence intervals, and LaTeX source packaging were verified using automated static consistency checks (`validate_and_package.py`).

References

[1] Mário S. Alvim, Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Anna Pazii. Local differential privacy on metric spaces: Optimizing the trade-off with utility. IEEE Computer Security Foundations Symposium (CSF), pages 262–276, 2018. URL https://doi.org/10.1109/CSF.2018.00026.

[2] Zhiyu An and Wan Du. Open problems in differentiable social choice, 2026. URL https://arxiv.org/abs/2602.03003.

[3] Prerna Arote, Adrian Saiz, Oriol Saguillo, and Lucianna Kiffer. Cryptographic security is not enough: Privacy gaps in the Renegade decentralized dark pool. arXiv preprint arXiv:2609.27100, 2026. URL https://arxiv.org/abs/2609.27100.

[4] Austin Bennett, Preston Vandervoort, Duc Le, and Mira Belenkiy. Concave is the new linear: The impossibility of anti-plutocratic DAO governance. arXiv preprint arXiv:2605.18990, 2026. URL https://arxiv.org/abs/2605.18990.

[5] Jacob Brodkey, Roberto Tamez, and Aaron Roth. Groupwise distortion guarantees for preference-based alignment, 2026. URL https://arxiv.org/abs/2610.05450.

[6] Aurélien Buchet, Soyong Kim, Tom Barbette, and Cristel Pelsser. Explicit QUIC proxies for server-side geo-blocking bypass. arXiv preprint arXiv:2610.06179, 2026. URL https://arxiv.org/abs/2610.06179.

[7] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. Bulletproofs: Short proofs for confidential transactions and more. IEEE Symposium on Security and Privacy (S&P), pages 315–334, 2018. URL https://doi.org/10.1109/SP.2018.00020.

[8] Jan Camenisch and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. Advances in Cryptology — EUROCRYPT, pages 93–118, 2001. URL https://doi.org/10.1007/3-540-44987-6_7.

[9] Alessandro Chiesa, Ziyi Guan, and Eylon Yıldız. Can AI oversight be zero knowledge? IACR Cryptology ePrint Archive, Paper 2026/2299, 2026. URL https://eprint.iacr.org/2026/2299.

[10] Vitaly Feldman, Audra McMillan, and Kunal Talwar. Counting on dummies: Privacy amplification with observable counts and truncated batches. arXiv preprint arXiv:2610.42139, 2026. URL https://arxiv.org/abs/2610.42139.

[11] Daniel Fett, Kristina Yasuda, and Brian Campbell. Selective disclosure for JWTs (SD-JWT). IETF RFC 9901, 2025. URL https://datatracker.ietf.org/doc/rfc9901/.

[12] Matteo Frigo and abhi shelat. Anonymous credentials from ECDSA (Longfellow-ZK). IACR Communications in Cryptology, 1(2), 2024. URL https://eprint.iacr.org/2024/2010.

[13] Paul Gölz, Nika Haghtalab, and Kunhe Yang. Distortion of AI alignment: Does preference optimization optimize for preferences? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 38, 2025. URL https://arxiv.org/abs/2505.23749.

[14] IEEE Standards Association / P7012 Working Group. Ieee std 7012-2025: Standard for machine readable personal privacy terms. IEEE Standards Association, 2025. URL https://standards.ieee.org/ieee/7012/7542/.

[15] Vladimir Kolesnikov, Ranjit Kumaresan, Mike Rosulek, and Ni Trieu. Efficient batched oblivious PRF with applications to private set intersection. ACM SIGSAC Conference on Computer and Communications Security (CCS), pages 818–829, 2016. URL https://doi.org/10.1145/2976749.2978381.

[16] Hao Li, Zhengkun Zhang, Gangqiang Hu, Zhen Zhang, Yude Gao, Dai Dai, and Jing Liu. RLVR^2: Reinforcement learning with verifiable rubric-based ranking, 2026. URL https://arxiv.org/abs/2609.23457.

[17] Junyi Li et al. Towards pareto-optimal tool-integrated agents with pareto ranking policy optimization (ParetoPO). International Conference on Machine Learning (ICML) / arXiv:2606.16111, 2026. URL https://arxiv.org/abs/2606.16111.

[18] Johnnatan Messias and Ayae Ide. Fairness in token delegation: Mitigating voting power concentration in DAOs. arXiv preprint arXiv:2510.05830, 2025. URL https://arxiv.org/abs/2510.05830.

[19] Kaisa Miettinen. Nonlinear multiobjective optimization. Kluwer Academic Publishers, 1999. URL https://doi.org/10.1007/978-1-4615-5563-6.

[20] Kazusato Oko, Annie Ulichney, Nika Haghtalab, and Han Bao. Distortion of AI alignment revisited: RLHF is a decent utilitarian aligner. In *International Conference on Machine Learning (ICML)*, 2026. URL https://arxiv.org/abs/2609.12651.

[21] Ralph E. Steuer and Eng-Ung Choo. An interactive weighted Tchebycheff procedure for multiple objective programming. Mathematical Programming, 26(3):326–344, 1983. URL https://doi.org/10.1007/BF02591870.

[22] E. Glen Weyl, Puja Ohlhaver, and Vitalik Buterin. Decentralized society: Finding Web3's soul. SSRN Working Paper 4105763, 2022. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4105763.

[23] Jay Yu, Austin Bennett, Billy Gao, and Rebecca Joseph. Evaluating voting design vulnerabilities for retroactive funding. arXiv preprint arXiv:2505.16068, 2025. URL https://arxiv.org/abs/2505.16068.