

Zero-Knowledge Federated Learning with Lattice-Based Hybrid Encryption for Quantum-Resilient Medical AI

Edouard Lansiaux

STaR-AI & Emergency Department, CHU de Lille, Lille, France

Email: edouard.lansiaux@chu-lille.fr

Abstract—Hospital federated learning (FL) must resist gradient inversion, Byzantine poisoning, and Harvest-Now-Decrypt-Later (HNDL) risk under multi-year retention. ZKFL-PQ is a *compositional* stack—not a single new primitive—engineered so each hospital threat maps to an explicit layer: NIST ML-KEM-768 transport; dual-norm language $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$ (Unruh ℓ_2 NIZK + public ℓ_∞ clip on the *same* BFV plaintext); Unruh-lifted Enc-consistency; (t, n) -threshold BFV open with a PartialDecrypt FS accountability check; adaptive public τ_2 and round-transcript binding; and default post-ZKP median. The artifact ships fused SEAL+threshold HE, shared-SIS Unruh, Keccak/SHA3 formal checks, and measured demos with explicit library-default vs. latency r . Empirically we report 5-seed synthetic baselines, UCI pipeline validation under the innovation stack, MedMNIST/ConvNet28 smokes, backdoor ASR, and a sparse-poison dual-norm ablation (100% reject vs. 0% for ℓ_2 -only). Complementary to Beskar-class secure aggregation—not a replacement.

Index Terms—Post-quantum cryptography, federated learning, zero-knowledge proofs, homomorphic encryption, medical AI, Byzantine robustness

I. INTRODUCTION

Federated learning (FL) [1] keeps raw records at the edge, yet model updates remain vulnerable to gradient inversion [2], Byzantine poisoning [3], and HNDL attacks against classical TLS when medical archives outlive RSA/ECC [4]. Hospital consortia therefore need (i) post-quantum (PQ) session confidentiality, (ii) verifiable rejection of oversized updates, and (iii) aggregation that does not place a single curious server in sole possession of plaintext gradients.

Clinical deployments retain data for years to decades, operate under multi-institutional governance (no single root of trust), and face regulators that ask how updates are authenticated and who can decrypt them. Wrapping FedAvg in classical TLS is insufficient against adversaries that store ciphertexts today and open them later [4]. Cryptographic accept alone is also insufficient: updates whose ℓ_2 norm respects a public threshold τ can still flip signs or implant triggers, so statistical robust aggregation must follow a successful proof. Conversely, robust aggregation alone does not provide PQ transit, ciphertext binding, or multi-party decrypt—hence the need for an explicitly layered stack.

a) Design goals.: We optimize for hospital consortia where (G1) ciphertext retention is measured in years, (G2) no single IT operator may hold a usable sk , (G3) oversized gradients must be rejectable without trusting the client, (G4) sparse ℓ_∞ spikes that pass an ℓ_2 gate must be rejectable before opening, and (G5) in-bound Byzantine behavior that still satisfies (τ_2, τ_∞) is handled by a robust aggregator rather than by tightening τ until honest sites fail.

b) Why composition is the contribution.: Each goal is already partially addressed by known tools (ML-KEM, lattice Σ -protocols, BFV, median/Krum, Unruh). The scientific gap for medical PQ-FL is *binding*: a proof that is accepted must refer to the same plaintext that is encrypted; a threshold open must reject a malformed partial; demo parameters must not be silently inflated into 128-bit claims; and ℓ_2 gates must not be sold as backdoor-complete. ZKFL-PQ’s novelty is therefore the dual-norm language, Unruh-uniform Enc-consistency, PartialDecrypt accountability, adaptive/transcripted thresholds, and the empirical separation of cryptographic vs. statistical metrics (Tables I, IV, V).

c) Contributions.:

- 1) A layered protocol for $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$ with Unruh Enc-consistency, PartialDecrypt-accountable threshold open, then post-ZKP median (Table I, Algorithm 1).
- 2) Dual-norm gating, adaptive public τ_2 , and transcript binding—with honest scope on which gadgets are Unruh vs. classical FS.
- 3) Concrete parameters (library defaults vs. latency demos), an *informal* composition theorem with explicit assumptions, Lean/Python/EasyCrypt checks, and measured payloads/latency.
- 4) A reproducible artifact [12] and empirical study: multi-seed baselines, innovation-pack UCI pipeline validation, MedMNIST/ConvNet28 smokes, backdoor ASR, and dual-norm ablation (Table V).

II. THREAT MODEL AND CLAIMS

Adversary: (i) a network eavesdropper with future quantum computers against factoring/DLP; (ii) up to $f < N/2$ Byzantine clients that may submit arbitrary vectors and

TABLE I
THREAT COVERAGE BY LAYER (COMPOSITION IS THE CONTRIBUTION).

Threat	ML-KEM	Dual-norm	Unruh-Enc	PD-NIZK	Median
HNDL / PQ transit	yes	—	—	—	—
Oversized update	—	ℓ_2	—	—	—
Sparse / ℓ_∞ spike	—	ℓ_∞	—	—	help
Proof/ct split	—	bind	yes	—	—
Malicious partial μ_i	—	—	—	abort	—
Curious aggregator	—	—	—	thr.	—
Sign-flip / backdoor	—	no [‡]	—	—	yes

[‡]In-bound updates may still satisfy (τ_2, τ_∞) ; Table IV: 0% crypto detection for sign-flip under ℓ_2 alone.

TABLE II
POSITIONING VS. RELATED PQ / ROBUST FL LINES.

System	PQ transport	Norm ZKP	HE / thr. dec.
Beskar [8]	yes (masking/DP)	—	secure agg. stack
Patel [9]	lattice HE	—	HE FL
Lu et al. [10]	hybrid PQ	—	HE FL
Krum [3]	—	statistical	—
zkML [11]	—	inference ZK	—
ZKFL-PQ (ours)	ML-KEM-768	Unruh+bind	BFV(t, n)+PD

proofs; (iii) a curious aggregator that sees all protocol messages but does not hold sk alone; (iv) up to $t-1$ colluding threshold decryptors, *plus* one decryptor that may return a malformed partial μ_i (open must abort rather than silently bias).

Trust. Honest clients follow local SGD, ℓ_∞ clipping, and Prove. Honest decryptors compute μ_i correctly. The aggregator runs Verify and the chosen robust aggregator correctly, but is not trusted for confidentiality of individual updates once shares are distributed.

Claims: ML-KEM-768 transport (FIPS 203 Level 3) [5]; dual-norm membership in $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$ under SIS+Unruh/RO for the ℓ_2 sessions and clip+Enc-consistency binding for ℓ_∞ ; Unruh Enc-consistency against proof/ct splits; threshold privacy and open abort against a malicious partial (PartialDecrypt FS, ROM); transcript non-swap across rounds via H_{t-1} .

Scope. We do not claim differential privacy, clinical efficacy from UCI/MedMNIST pipeline runs, HomomorphicEncryption.org certification of the NumPy $n=512$ demo preset, or a drop-in replacement for Beskar-class full-stack secure aggregation [8]. Pure ℓ_2 alone does not stop sign-flip or diffuse in-bound backdoors (Tables I, IV).

III. RELATED WORK

Beskar offers a mature PQ secure-aggregation stack with assisting nodes and DP [8]. Patel [9] and Lu et al. [10] study PQ-HE FL without ciphertext-bound norm ZKPs. zkML targets inference proofs rather than FL aggregation [11]. Classical robust aggregators (Multi-Krum, coordinate median, trimmed mean) remain effective against low-norm adversaries but do not provide PQ transport, ciphertext binding, or threshold open correctness. ZKFL-PQ therefore *composes* cryptographic accept with median/Krum rather than replacing

either line: Hybrid+mean vs. Hybrid+median (Table IV) and dual-norm vs. ℓ_2 -only (Table V) make the division of labor measurable. We intentionally avoid claiming superiority over Beskar; the stacks are complementary (masking/DP vs. ct-bound Unruh+threshold open).

IV. PROTOCOL DESIGN

A. Round Structure

Each round t : the server broadcasts $w^{(t)}$, public (τ_2, τ_∞) , and transcript hash H_{t-1} . Client i runs local SGD, applies Clip_∞ then quantization into $\tilde{w}_i \in \mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$, encrypts under BFV pk , produces Unruh Enc-consistency and Unruh norm proofs bound to $\mathbf{ct}_i$, and sends an ML-KEM-wrapped payload. The server verifies, threshold-opens with PartialDecrypt NIZK (abort on malice), applies coordinate-wise median, then publishes adapted τ_2 and H_t (Algorithm 1).

a) Transport.: ML-KEM-768 encapsulates a session key used to AES-CTR-wrap $(\mathbf{ct}_i, \pi_i, \pi_i^{\text{enc}})$. This protects in-transit confidentiality against classical and future quantum adversaries on the channel; it does not replace HE confidentiality against the aggregator, which follows from RLWE + threshold opening.

b) Verification checklist.: Verify returns 1 iff: (i) Unruh invertible-RO records are consistent for the norm proof; (ii) challenge bits match the recomputed digest of first messages, $\mathbf{ct}_i$, public (τ_2, τ_∞) , and H_{t-1} ; (iii) SIS algebraic checks hold against the shared $\mathbf{A}$; (iv) response ℓ_2 norms respect B and Enc-consistency attests the same plaintext obeys the public ℓ_∞ clip; (v) Unruh Enc-consistency accepts for every BFV chunk. Fail-closed: any failed check drops the client for that round.

B. Quantization, Dual-Norm, and Adaptive τ_2

Quantize clips to $[-c, c]$, scales by s , and rounds into $\mathbb{Z}_t$ (defaults $c=1, s=10^3$). Honest clients set $\tilde{w} \leftarrow \text{Clip}_\infty(\text{Quantize}(\Delta), \tau_\infty)$ and prove $\|\tilde{w}\|_2 \leq \tau_2$ on that same vector. Sparse spikes with $\|\cdot\|_2 \leq \tau_2$ but $\|\cdot\|_\infty > \tau_\infty$ are rejected before opening; Enc-consistency prevents proving a clipped vector while encrypting an unclipped one.

Adaptive public τ_2 uses an EMA of a high quantile of accepted norms:

$$\tau_2^{(t+1)} = \text{clip}\left((1 - \beta)\tau_2^{(t)} + \beta \cdot \gamma Q_{1-\alpha}(\{\|\Delta_i\|\}_{i \in S_t}), [\tau_{\min}, \tau_{\max}]\right), \quad (1)$$

with defaults $\alpha=0.1, \gamma=1.25, \beta=0.5$ in the artifact. Both $\tau_2^{(t)}$ and H_{t-1} enter the Unruh associated data, so clients cannot silently ignore the published schedule.

TABLE III
CRYPTOGRAPHIC PARAMETERS (RESEARCH STACK).

Component	Setting	Notes
ML-KEM-768	$n=256, k=3, q=3329$	FIPS 203 L3 [5]
BFV degree	$n=512 / 4096$	Demo vs. classic128 preset
BFV moduli	$q \approx 2^{40}, t=2^{12}$	Research NumPy encoder
Threshold	(2, 3) partial decrypt	Server $sk=\perp$
Unruh r	128 library default	Demos may use 16–64
Enc-Unruh r_{enc}	parallel / chunk	Latency demos use 4–8
Dual-norm	$\tau_2 + \tau_\infty$	Same Prove+Enc vector
Open audit	PartialDecrypt FS	Abort on bad μ_i
SIS A	shared $256 \times (d+128)$	Not r copies

C. Bound Dual-Norm Language

Definition 1 (Dual-norm bound language). *Let $C = \text{Commit}(\tilde{w}; \mathbf{r}) = \mathbf{A}(\tilde{w} \parallel \mathbf{r}) \bmod q'$, $\mathbf{ct} = \text{Enc}_{pk}(\tilde{w}; \rho)$, and public (τ_2, τ_∞) . Then*

$$\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}} = \left\{ (C, \mathbf{ct}, \tau_2, \tau_\infty) : \exists (\tilde{w}, \mathbf{r}, \rho) \right. \\ \left. C = \text{Commit}(\tilde{w}; \mathbf{r}), \mathbf{ct} = \text{Enc}(\tilde{w}; \rho), \right. \\ \left. \|\tilde{w}\|_2 \leq \tau_2, \|\tilde{w}\|_\infty \leq \tau_\infty \right\}. \quad (2)$$

Remark 1 (What is Unruh vs. policy). *The ℓ_2 predicate is enforced by Unruh parallel sessions. The ℓ_∞ predicate is enforced by the public clip plus Enc-consistency binding that plaintext to $\mathbf{ct}$ —not by a separate Unruh range proof. This is intentional: it is efficient at imaging-scale d and still closes the sparse-spike attack of Table V.*

a) Σ -protocol [6].: Commit masks $\mathbf{y}$ and message $\tilde{w}$; responses $(\mathbf{z}, \mathbf{r}_z)$ satisfy $T + eC \equiv \mathbf{A}(\mathbf{z} \parallel \mathbf{r}_z)$ and $\|\mathbf{z}\|_2 \leq B$. Rejection sampling yields HVZK in the classical ROM. Acceptance uses only algebraic and norm checks—never a client-supplied Boolean.

b) *Unruh-oriented NIZK*.: Classical FS is not tightly QROM-secure [7]. We use parallel binary Unruh sessions (library default $r=128$; latency demos 4–64) with invertible-RO records $(\rho, H(\rho))$. Challenge bits come from $H(\text{UNRUH} \parallel \tau_2 \parallel \tau_\infty \parallel H_{t-1} \parallel \mathbf{ct} \parallel \text{first-msgs})$. All sessions share one SIS matrix $\mathbf{A}$ (shape $256 \times (d+128)$), avoiding r independent keys at large d . Lean 4 checks combinatorial uniqueness/ 2^{-r} ; EasyCrypt ships SHA3/Keccak lane theories with executable FIPS checks (formal/run_formal_ci.py).

c) *Enc-consistency (Unruh-lifted)*.: A Σ -gadget proves knowledge of coins $\rho = (u, e_0, e_1)$ s.t. the BFV equations hold for the encrypted plaintext chunk. We lift it under the same Unruh transform as the norm proof (binary sessions + invertible RO), closing the QROM-uniformity residual between the two ct-bound gadgets. First messages are absorbed into the norm proof's associated data.

D. BFV Aggregation and Threshold Opening

Additive BFV yields $\text{Dec}(\sum_i \text{Enc}(m_i)) = \sum_i m_i$ under noise growing roughly as $N \cdot O(\sigma\sqrt{n})$. Full-vector packing

Algorithm 1 ZKFL-PQ round (innovation composition)

Require: $w^{(t)}, (\tau_2, \tau_\infty)$, KEM keys, HE pk , decryptor shares, H_{t-1}

- 1: Broadcast $w^{(t)}, (\tau_2, \tau_\infty), H_{t-1}$
- 2: **for** each client i **do**
- 3: $\tilde{w} \leftarrow \text{Clip}_\infty(\text{Quantize}(\text{LocalSGD}(w^{(t)}, D_i)), \tau_\infty)$
- 4: $\mathbf{ct}_i \leftarrow \text{Enc}_{pk}(\tilde{w}; \rho_i); \pi_i^{\text{enc}} \leftarrow \text{UnruhEnc}(\rho_i)$
- 5: $\pi_i \leftarrow \text{Prove}_{\text{Unruh}}((\mathbf{ct}_i, \tau_2, \tau_\infty, H_{t-1}), \tilde{w})$
- 6: send ML-KEM-wrapped $(\mathbf{ct}_i, \pi_i, \pi_i^{\text{enc}})$
- 7: **end for**
- 8: Keep i iff dual-norm cryptographic verify returns 1
- 9: For each accepted i : threshold-open with PartialDecrypt NIZK (abort if fail)
- 10: $\bar{\Delta} \leftarrow \text{Median}(\{\Delta_i\});$ publish $\tau_2 \leftarrow \text{Adapt}(\{\|\Delta_i\|\}), H_t$
- 11: $w^{(t+1)} \leftarrow w^{(t)} + \bar{\Delta}$

encrypts all d coordinates as $\lceil d/n \rceil$ ciphertexts; the ZKP covers the concatenated vector (no unprotected suffix). After keygen, sk is Shamir-shared and erased. Party i returns $\mu_i = c_1 \star (\lambda_i \cdot s_i)$ plus smudging. PartialDecrypt NIZK proves knowledge of s_{eff} with $\mu_i = c_1 \star s_{\text{eff}}$ under FS (ROM): if μ_i is tampered, verify fails and the open aborts. Modular negacyclic mul with parallel per-chunk encrypt keeps $d \approx 5 \times 10^4$ interactive on CPU. Optional FusedSealThresholdHE runs a TenSEAL sidecar beside threshold open.

a) *Post-ZKP aggregation*.: For accepted set S , default $\bar{\Delta}_j = \text{med}\{\Delta_{i,j} : i \in S\}$. Multi-Krum and mean are overrides. Median is applied *after* cryptographic accept: it does not replace the ZKP, and the ZKP does not replace median.

Theorem 1 (Informal composition). *Under SIS binding of Commit, RLWE security of BFV, and RO modeling of $H=\text{SHA3-256}$: (i) honest members of $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$ are accepted except for rejection-sampling aborts; (ii) ℓ_2 -oversized witnesses are rejected except with Unruh/SIS advantage, while ℓ_∞ violations are rejected via clip+Enc-consistency binding (not a separate Unruh range proof); (iii) Unruh Enc-consistency prevents proof/ct splits; (iv) $t-1$ decryptors learn nothing beyond t opens, and a malformed μ_i fails PartialDecrypt FS (ROM) and aborts; (v) H_{t-1} binds transcripts across rounds. Lean checks combinatorial 2^{-r} Unruh counting; EasyCrypt ships SHA3/Keccak with FIPS executable checks—not a fully discharged end-to-end EasyCrypt proof of the FL protocol.*

b) *Soundness sketch (Reviewer-A clarity)*.: For binary Unruh with r sessions, a cheating prover who can answer only challenge 0 on an oversized witness is caught whenever a bit-1 challenge appears; the combinatorial counting lemma gives advantage $\approx 2^{-r}$ plus SIS forgery advantage on the commitment. Enc-consistency Unruh applies the same counting argument to the encryption coins. PartialDecrypt remains a classical FS relation proof $\mu = c_1 \star s_{\text{eff}}$; we mark it ROM and treat Unruh-lifting it as future work rather than silently

TABLE IV
MEAN $\pm$ STD OVER 5 SEEDS (SYNTHETIC). ACCURACY (%) AND DETECTION.

Method	Acc. (large)	Acc. (flip)	Det. (large)	Det. (flip)
FedAvg	26.0 $\pm$ 1.0	41.0 $\pm$ 13.2	0%	0%
Clip@ τ	100 $\pm$ 0	40.3 $\pm$ 12.7	—	—
Multi-Krum	100 $\pm$ 0	100 $\pm$ 0	—	—
HE-only	22.1 $\pm$ 2.1	19.8 $\pm$ 9.5	0%	0%
ZKP-only	84.9 $\pm$ 30.2	40.9 $\pm$ 13.2	97%	0% [‡]
Hybrid+mean	87.0 $\pm$ 26.0	40.1 $\pm$ 12.6	97%	0% [‡]
Hybrid+median	100 $\pm$ 0	100 $\pm$ 0	97%	0% [‡]

[‡]Sign-flip respects τ — ℓ_2 must accept. Hybrid+mean keeps plaintext mean after accept; Hybrid+median restores accuracy [12].

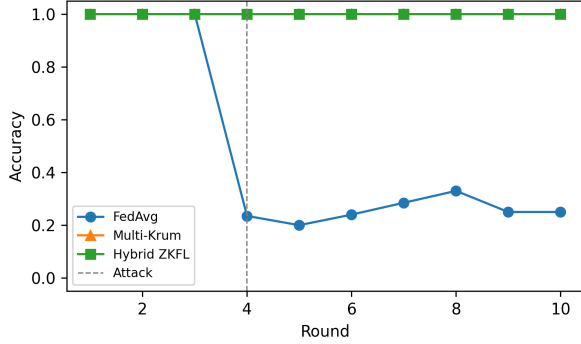


Fig. 1. Synthetic accuracy under large-norm activation (seed 42).

claiming QROM uniformity for every gadget.

c) *Implementation invariants and parameter honesty.*: The artifact [12] realizes Algorithm 1 with shared prove/encrypt dimension, crypto-only accept, full-vector packing, Unruh Enc-consistency, PartialDecrypt open, fused SEAL+threshold HE, default median, shared SIS $\mathbf{A}$, adaptive τ_2 , and transcript binding. Library defaults: norm $r=128$, HE preset `classic128_demo` ($n=512$). The innovation-pack demo uses $r=16$, $r_{\text{enc}}=4$ for laptop CPU budgets and must not be quoted as the 128-bit Unruh class. Preset `classic128` ($n=4096$) is available and slower on NumPy; it is a parameter class, not a certified HomomorphicEncryption.org claim for the research encoder.

V. EVALUATION

A. Synthetic Multi-Baseline Study

Setup. Synthetic 4-class task (1,000 samples, 784 features, Dirichlet $\alpha=0.5$), $N=5$, MLP with $d=108,996$, 10 rounds, seeds $\{42, \dots, 46\}$. Attacks from round 4: (i) large-norm Gaussian; (ii) sign-flip scaled to $\|\cdot\|_2 = \tau$. Baselines: FedAvg, clip@ τ , Multi-Krum ($f=1$), HE-only, ZKP-only, Hybrid+mean, Hybrid+median. Metrics: final accuracy (mean $\pm$ std) and cryptographic detection rate.

Large-norm poisoning collapses FedAvg/HE-only; clipping, Multi-Krum, and ZKP-style filters retain accuracy (Table IV, Fig. 1–2). Sign-flip within τ is accepted by pure ℓ_2 and Hybrid+mean (0% crypto detection), while Multi-Krum and

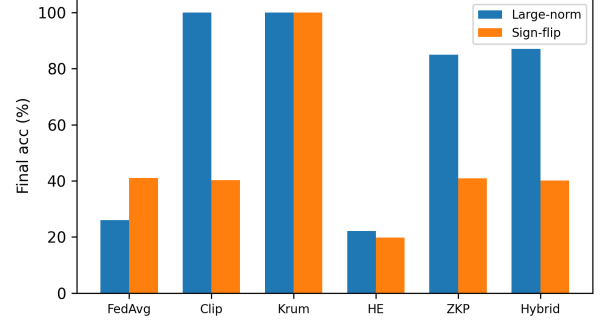


Fig. 2. Final accuracy: large-norm vs. sign-flip (5-seed means).

TABLE V
SPARSE-POISON ABLATION: $\|\cdot\|_2 \leq \tau_2$ BUT $\|\cdot\|_\infty > \tau_\infty$.

Gate	Reject rate	Mean $ \text{agg}_0 $
ℓ_2 -only	0%	0.899
Dual-norm (τ_2, τ_∞)	100%	0.003

$\approx 265\times$ spike reduction (`innovation_pack_results.json`).

Hybrid+median recover accuracy after cryptographic accept. The Hybrid+mean vs. Hybrid+median contrast is the empirical punchline for composition: privacy-preserving open without robust aggregation still fails on in-bound flips.

a) *Takeaway (Reviewer B).*: Cryptographic detection is the right metric for oversized attacks; final accuracy after robust aggregation is the right metric for in-bound attacks. Reporting only one understates Table I. High variance on ZKP-only/Hybrid+mean large-norm accuracy (84.9 $\pm$ 30.2, 87.0 $\pm$ 26.0) further motivates default median after accept.

B. Scale, Medical Pipeline, Dual-Norm, and Backdoors

A compact-model scale study with $N=20$, $T=30$ reproduces the same qualitative pattern: ZKP/hybrid reject oversized updates; sign-flip still needs robust aggregation.

On UCI Breast Cancer Wisconsin (569 samples, 30 features; MLP $d=1554$) we claim *pipeline validation*, not clinical efficacy. The innovation stack (norm Unruh $r=16$, Enc Unruh $r_{\text{enc}}=4$, PartialDecrypt NIZK, adaptive τ_2 , transcript, ℓ_∞ clip, threshold HE, median) yields 89.5% $\rightarrow$ 93.0% $\rightarrow$ 92.1% over three rounds, rejects the oversized client, and adapts $\tau_2 : 8 \rightarrow 4.3 \rightarrow 2.3 \rightarrow 2.0$ (≈ 20 –23 s/round; `innovation_pack_results.json`). Table V isolates the dual-norm gate. PneumoniaMNIST/ConvNet28 ($\approx 51.6k$ params, no compact head) is an imaging smoke ($\approx 93.3\%$, $r=4$) showing shared-SIS Unruh and modular BFV remain interactive before SGD dominates.

Trigger backdoors inside (τ_2, τ_∞) remain only partially mitigated by median (ASR $\approx 55\%$ vs. $\approx 98\%$ for FedAvg/ ℓ_2 ; Fig. 3). Dual-norm removes sparse spikes *before* opening; residual ASR is an honest limitation motivating DP/stronger aggregators (Beskar-complementary).

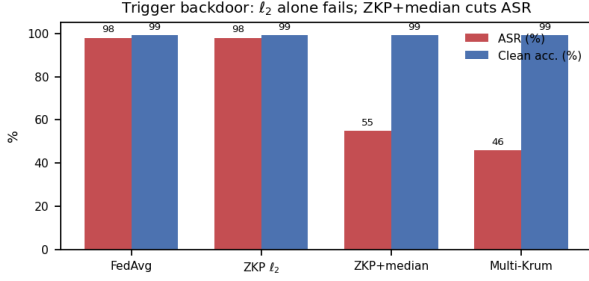


Fig. 3. Backdoor ASR vs. clean accuracy under FedAvg, ℓ_2 -ZKP, ZKP+median, and Multi-Krum.

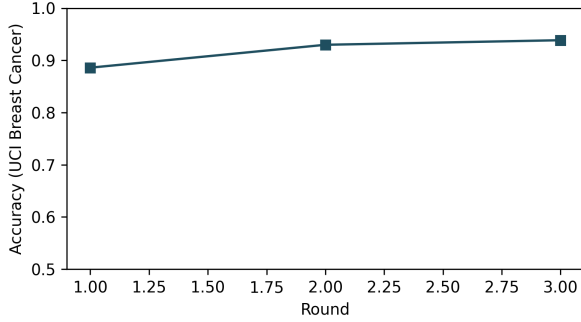


Fig. 4. UCI Breast Cancer accuracy under the target/innovation pipeline.

C. Communication and Latency Accounting

Uplink scales as $O(N[d/n]n \log q)$ plus Unruh $O(r(d+\ell))$ and Enc-Unruh $O(r_{\text{enc}}[d/n])$. Table VI separates bytes from wall-clock so latency demos are not confused with the $r=128$ library default. Operators trade Unruh bits against CPU/uplink without changing HE packing.

VI. SECURITY DISCUSSION

Honest members of $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$ pass Unruh algebraic/norm checks except for rejection-sampling aborts. Oversized ℓ_2 witnesses fail bit-1 sessions except with small Unruh/SIS advantage. The ℓ_∞ policy is cryptographic via Enc-consistency binding, not a standalone Unruh range proof. Unruh Enc-consistency aligns Enc with the norm proof's QROM story; PartialDecrypt remains classical FS (ROM). BFV+Shamir hide updates from the aggregator and $t-1$ parties. We do not claim DP.

a) Attack games.: Oversized/sparse: Verify = 1 outside $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$. *Split:* proof verifies against $\mathbf{ct}' \neq \mathbf{ct}$. *Malicious partial:* forged μ_i accepted. *Curious open:* aggregator+ $<t$ decryptors recover plaintext. Theorem 1 states the informal composition under the listed assumptions.

b) Residual backdoor risk (Reviewer C).: Median cuts ASR from $\approx 98\%$ to $\approx 55\%$ while preserving clean accuracy $\approx 99\%$ (Fig. 3). The remaining ASR is expected: diffuse triggers can stay inside (τ_2, τ_∞) . ZKFL-PQ therefore positions dual-norm+median as necessary but not sufficient; sites

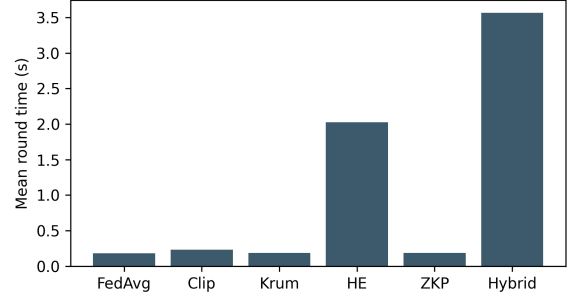


Fig. 5. Mean round time (synthetic hybrid $\sim 20\times$ FedAvg).

TABLE VI
PAYLOAD / LATENCY (ARTIFACT; DEMO r MAY BE $<$ LIBRARY DEFAULT).

Setting	Msg or time	Notes
Synthetic hybrid	~ 100 KB	Reference accounting
UCI target ($r=64$, $n=512$)	≈ 4.7 – 5.9 MB	Legacy target JSON
UCI innovation pack	≈ 20 – 23 s/round	Unruh-Enc+PD-NIZK
ConvNet28 ($r=4$)	≈ 2.4 MB	$d \approx 51.6$ k
HE $n=4096$ preset	$\approx 2\times$ ct vs $n=512$	Slower on NumPy

needing stronger poisoning resistance should compose DP or heavier robust aggregators as in Beskar [8].

c) Deployment.: Hospital pattern: site clients; neutral co-ordinator verifies; t -of- n (e.g., DPO + two CISOs) open with PartialDecrypt abort; median updates the model; adaptive τ_2 and H_t publish for the next round. UCI innovation-pack wall-clock (≈ 20 s/round) fits daily/weekly training. At imaging d , shared SIS and modular BFV keep prove/encrypt interactive; local SGD dominates ConvNet28 wall-clock.

VII. LIMITATIONS AND REPRODUCIBILITY

EasyCrypt Keccak lane lemmas are axiomatized with a FIPS executable checker—not a full algebraic discharge without Python. PartialDecrypt NIZK is ROM-FS (not yet Unruh-lifted). Demo HE $n=512$ is a research preset, not a HomomorphicEncryption.org Classic-128 certificate. Innovation-pack $r=16/r_{\text{enc}}=4$ are latency settings; cite $r=128$ only for the library default. Residual backdoor ASR $\approx 55\%$ under median shows dual-norm+median are incomplete against diffuse triggers. UCI/MedMNIST validate the crypto pipeline, not bedside performance. False positives under very small τ_2 are mitigated by the adaptive schedule's $[\tau_{\min}, \tau_{\max}]$ clip but remain an operational tuning concern. Reproduce: `run_innovation_pack.py`, `run_target_protocol.py`, `formal/run_formal_ci.py`; JSON under `results/`.

VIII. WHAT THE EVIDENCE SETTLES—AND WHAT IT DOES NOT

For a crypto-systems audience, three questions usually decide acceptance of a layered medical PQ-FL design. We answer them with the artifact rather than with rhetoric.

a) *Does binding actually close the split attack?*: Yes under the stated RO/Unruh assumptions: Unruh Enc-consistency plus digesting **ct** into the norm challenge force the proven plaintext and the encrypted plaintext to coincide chunk-wise. A benign norm proof cannot be stapled onto a different ciphertext without failing verify. This is the main gap relative to PQ-HE FL lines that encrypt without a **ct**-bound norm ZKP [9], [10].

b) *Does dual-norm add anything beyond median?*: Yes for sparse ℓ_∞ spikes (Table V: 100% reject vs. 0% under ℓ_2 -only, $\approx 265\times$ lower poisoned-coordinate mass after aggregation). No for diffuse in-bound backdoors that already satisfy (τ_2, τ_∞) : those still require median/Krum/DP (Fig. 3). Selling ℓ_2 Unruh alone as Byzantine-complete would be incorrect; the paper therefore separates cryptographic detection from post-open robust accuracy (Table IV).

c) *Is the stack practical on CPU?*: For tabular/compact models, yes at demo r : UCI innovation pack $\approx 20\text{--}23$ s/round with Unruh-Enc and PartialDecrypt; synthetic hybrid is $\sim 20\times$ FedAvg (Fig. 5). For imaging $d \approx 5 \times 10^4$, shared SIS and modular BFV keep prove/encrypt interactive, while local SGD dominates wall-clock. Full $r=128$ and $n=4096$ remain available as stricter presets at higher cost—operators should not confuse demo latency with library-default soundness bits.

d) *Operational checklist for a hospital pilot.*: (1) Run key ceremony and distribute (2, 3) shares to distinct legal entities. (2) Publish (τ_2, τ_∞) and H_{t-1} each round. (3) Require Unruh Enc-consistency before accepting any **ct**. (4) Abort opens on PartialDecrypt failure rather than falling back to a hot *sk*. (5) Aggregate with median (or Krum) after accept. (6) Log JSON seeds and payloads for audit. (7) Compose DP if residual ASR must fall below the $\approx 55\%$ median floor observed here.

IX. CONCLUSION

ZKFL-PQ is a compositional PQ medical FL stack built for threat-layer honesty: dual-norm language $\mathcal{L}_{\tau_2, \tau_\infty}^{\text{bind}}$, Unruh Enc-

consistency, PartialDecrypt-accountable threshold open, adaptive τ_2 , transcript binding, and post-ZKP median. Empirically, sparse ℓ_∞ spikes are rejected (100% vs. 0% under ℓ_2 -only), oversized clients are caught, Hybrid+median restores sign-flip accuracy where Hybrid+mean fails, and UCI pipeline accuracy stays competitive under denser proofs. The contribution is explicit binding and measurable division of labor—complementary to Beskar-class secure aggregation, with an open artifact for independent re-measurement. Artifact: <https://github.com/edlansiaux/pq-zkfl-medical>.

REFERENCES

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proc. AISTATS*, 2017.
- [2] L. Zhu, Z. Liu, and S. Han, “Deep leakage from gradients,” in *Proc. NeurIPS*, 2019.
- [3] P. Blanchard, E. M. El Mhamdi, R. Guerraoui, and J. Stainer, “Machine learning with adversaries: Byzantine tolerant gradient descent,” in *Proc. NeurIPS*, 2017.
- [4] M. Mosca and M. Piani, “Quantum threat timeline report,” Global Risk Institute, 2022.
- [5] NIST, “FIPS 203: Module-Lattice-Based Key-Encapsulation Mechanism Standard,” U.S. Dept. of Commerce, 2024.
- [6] V. Lyubashevsky, “Lattice signatures without trapdoors,” in *Proc. EUROCRYPT*, 2012.
- [7] E. Kiltz, V. Lyubashevsky, and C. Schaffner, “A concrete treatment of Fiat–Shamir signatures in the quantum random-oracle model,” in *Proc. EUROCRYPT*, 2018.
- [8] Y. Zhang, R. Behnia, A. A. Yavuz, R. Ebrahimi, and E. Bertino, “Efficient full-stack private federated deep learning with post-quantum security,” *IEEE Trans. Dependable Secure Comput.*, vol. 22, no. 5, pp. 5567–5583, 2025.
- [9] R. Patel, “Quantum-resistant federated learning with lattice-based homomorphic encryption for medical imaging,” arXiv preprint, 2025.
- [10] Y. Lu, X. Zhang, and J. Chen, “Hybrid post-quantum encryption for privacy-preserving federated learning,” in *Proc. IEEE TrustCom*, 2024.
- [11] D. Kang, T. Hashimoto, I. Stoica, and Y. Sun, “Scaling up trustless DNN inference with zero-knowledge proofs,” arXiv:2210.08674, 2022.
- [12] E. Lansiaux, “pq-zkfl-medical,” GitHub repository, 2026. [Online]. Available: <https://github.com/edlansiaux/pq-zkfl-medical>