Roster or Reputation?

Attributing Sybil Resistance to Mechanisms in Secure Clustered Wireless Sensor Networks: A Simulation Study

Saidi Adetunji Adegoke[1] · K. A. Abdulsalam[1] · N. A. Azeez[2]
[1] Dept. of Electrical and Electronics Engineering · [2] Dept. of Computer Sciences
University of Lagos, Lagos, Nigeria

THE PROBLEM

The convention observed

Secure clustering protocols bundle defensive mechanisms, then report **one** detection or admission figure for the combination.

MECHANISM 1

Registration roster

A permissioned authority set: only identities enrolled at deployment may lead.

MECHANISM 2

Behavioural trust

A trust score that bars nodes whose record falls below a threshold.

MECHANISM 3

Consensus quorum

A ledger (usually Proof-of-Authority) whose quorum must approve each block.

Why it matters: clustered networks watch fields, and also **bridges, pipelines and water networks**, where one captured cluster head can suppress or falsify its whole cluster’s readings.

THE PROBLEM

Three mechanisms, three different observables

They are **not three implementations of one defence**. Each observes something different, so each reaches a different adversary population.

Mechanism	Observes	How the test behaves
Roster (M_R)	Provenance: was this identity enrolled?	No evidence needed; decides in one round
Veto (M_B)	Conduct: is its trust record above τ?	Needs evidence; decides after k rounds
Quorum (M_Q)	Agreement: do the authorities concur?	Block commits on [N_c/2] votes; fails only in a majority-compromised cluster

The timing of each test follows from what it observes.

THE CLAIM

> A behavioural mechanism cannot defend against an adversary who can discard and re-mint identities, and a provenance mechanism cannot defend against an adversary who has compromised a legitimately enrolled node.

Douceur (2002) states the first half analytically.

What the literature lacks is not the theorem but the **measurement**: an experiment that crosses the mechanisms, grants the adversary the capability the theorem assumes, and counts the two populations separately.

THE MODEL

Two adversary populations

H

Compromised hardware

Physical nodes, enrolled at deployment under legitimate identities, whose behaviour has since turned adversarial.

Bound to a provenance record. Not renewable.

F

Forged identities

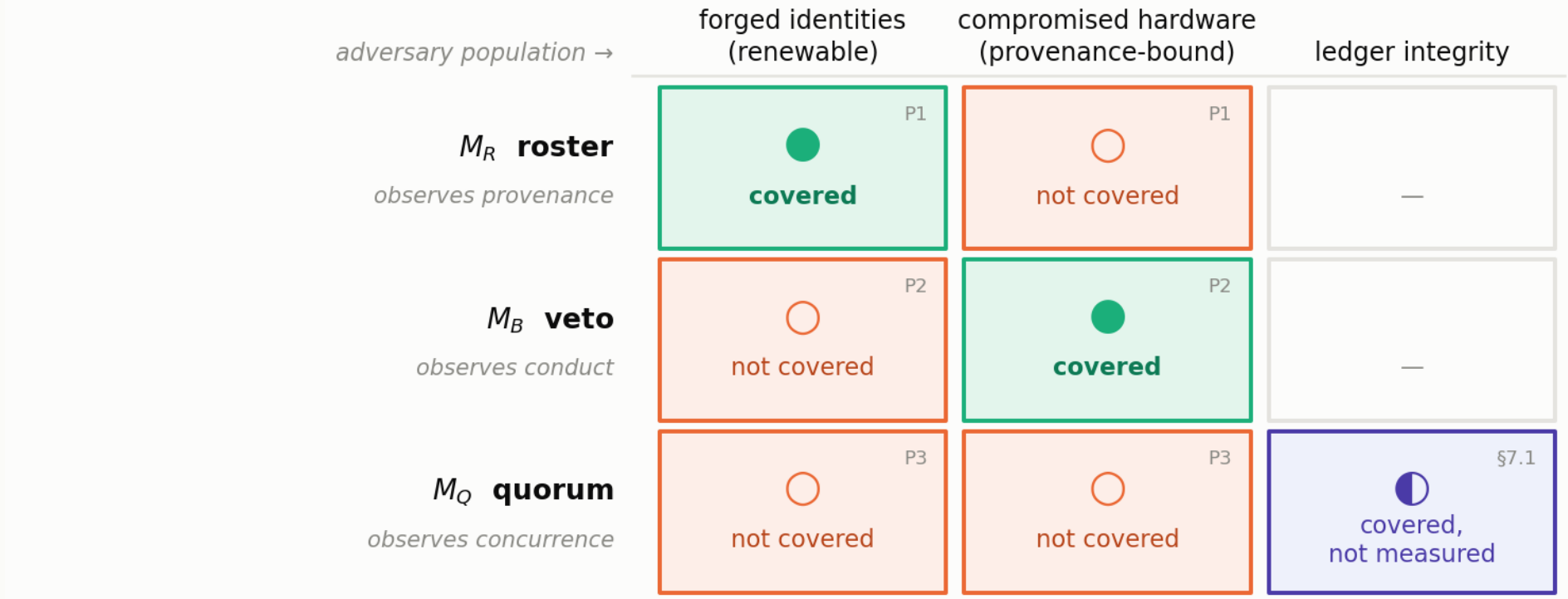
Not devices: the attacker's single radio announcing itself under an additional name, with a **clean behavioural history**.

Unbound. Freely renewable.

Every attribution result follows from that single asymmetry.

THE MODEL

The coverage relation



Each mechanism is characterised by what it observes; that observable decides which population it can reach.

THE MODEL

Three propositions

P1 · ROSTER $\succ$ F

Complete against minting

A freshly minted name is off-roster exactly as a discarded one was. Total, evidence-free, and **invariant to the minting rate**.

P2 · VETO $\not\succ$ F

Incomplete under churn

A re-minting adversary stays eligible a fraction $k/(k+\delta)$ of rounds, with one delay $\boldsymbol{\delta = 1 + d}$. The converse holds: Veto $\succ$ H, since hardware cannot whitewash itself.

P3 · QUORUM $\not\succ$ F, H

Orthogonal

A quorum is a **block-acceptance** test, not an admission test. It defends a third channel: ledger integrity.

d = any additional re-announcement delay. A replacement identity can stand for election no earlier than the next round, hence the 1.

THE MODEL

Measurement framework: five instruments

1. **Population-separated metrics:** forged-CH share, compromised-hardware share, and the aggregate
2. **Identity churn as the default threat model:** one-shot minting is only the ablation
3. **Factorial mechanism ablation:** roster × veto, 2×2, across attacker fractions
4. **Quorum activity reporting:** block-rejection rate, kept separate from admission
5. **Sweeping both time constants:** the defender’s k and the adversary’s delay d

Each instrument maps one-to-one onto a reporting requirement, R1 to R5 (slide 22).

THE MODEL

Experimental setup

HOST AND BASELINES

STITCH: Interval Type-2 fuzzy clustering with a Proof-of-Authority authority set, fully specified (trust update as Eq. 1; every parameter in the supplement). *The harness, not the contribution.*

Baselines: E-FLZSEPFCH (unsecured) · EA-PoA (roster + quorum, no veto)

SCALE

N = 100 in 200×200 m², attackers 10 / 20 / 30 %, s = 3 forged identities each

Replicated on **HEED** and **Tao IT2-TSK**, and at N = 200, 300, 500

3,960

runs, 20 seeds each

STATISTICS

Shares: seed-level **bootstrap 95 % CIs** and **Wilcoxon** signed-rank tests. Half-life: Student-t CIs. Code and data open: doi 10.5281/zenodo.24244917

RESULTS

Read on the aggregate, PoA looks like theatre

Under Sybil, PoA is **worse** on the aggregate: 0.204 against 0.179, outside both intervals. **That conclusion is wrong.**

Protocol (Sybil, 20 % compromised)	Malicious-CH admission	Forged-CH share
E-FLZSEPFCH (unsecured)	0.179 ± 0.018	0.123 ± 0.011
EA-PoA (roster + quorum)	0.204 ± 0.019	0.000 ± 0.000
STITCH	0.001 ± 0.000	0.000 ± 0.000

The forged-CH column, which the conventional metric never separates out, shows PoA holding fabricated identities at **exactly zero**.

20 seeds, mean ± 95 % CI.

RESULTS

The quorum is active, and still does not matter

THE GATE FIRES

0.107

EA-PoA block-rejection rate at 30 % compromised, up from 0.006 at 10 %

ADMISSION IS UNTOUCHED

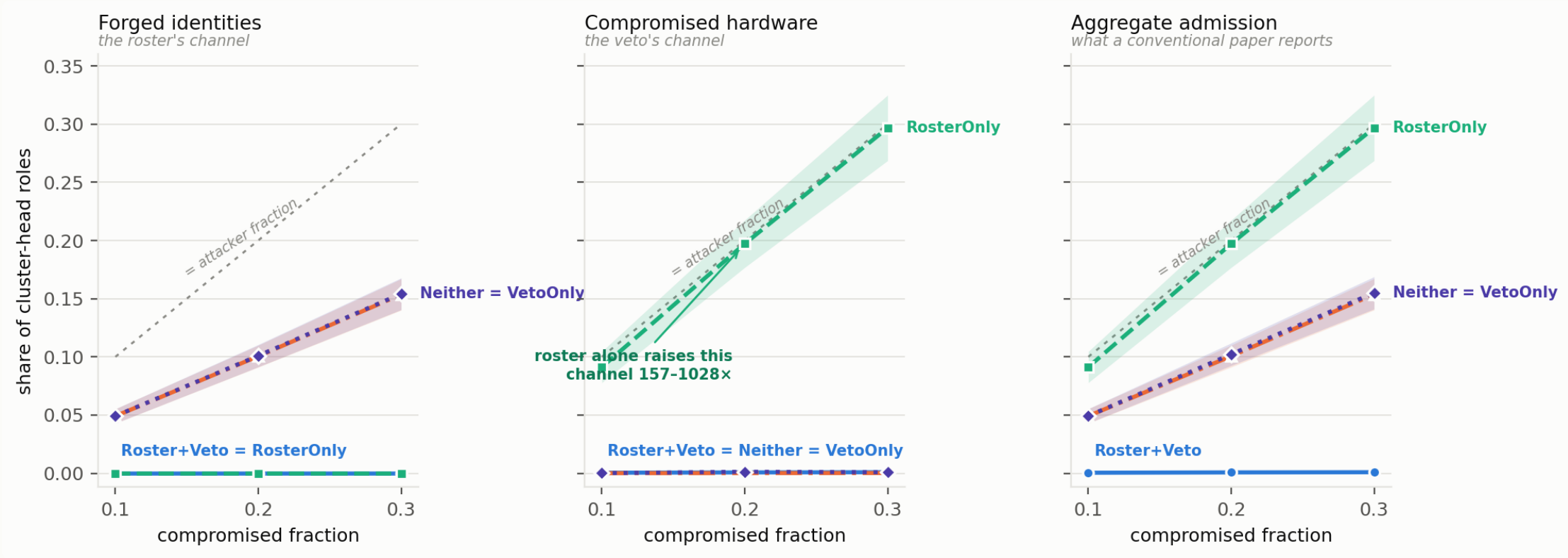
0.204

aggregate admission at 20 % attackers (0.101 at 10 %, 0.308 at 30 %): it tracks the attacker fraction; delivery unmoved at 0.985

> A consensus layer that does not gate election cannot change which nodes are elected.

Values with CIs: Supplementary Table S3(a).

Separate the populations



Left: the roster’s channel. Centre: the veto’s channel. Right: the only one the field reports. 20 seeds, 95 % CI bands.

RESULTS

The factorial, decomposed

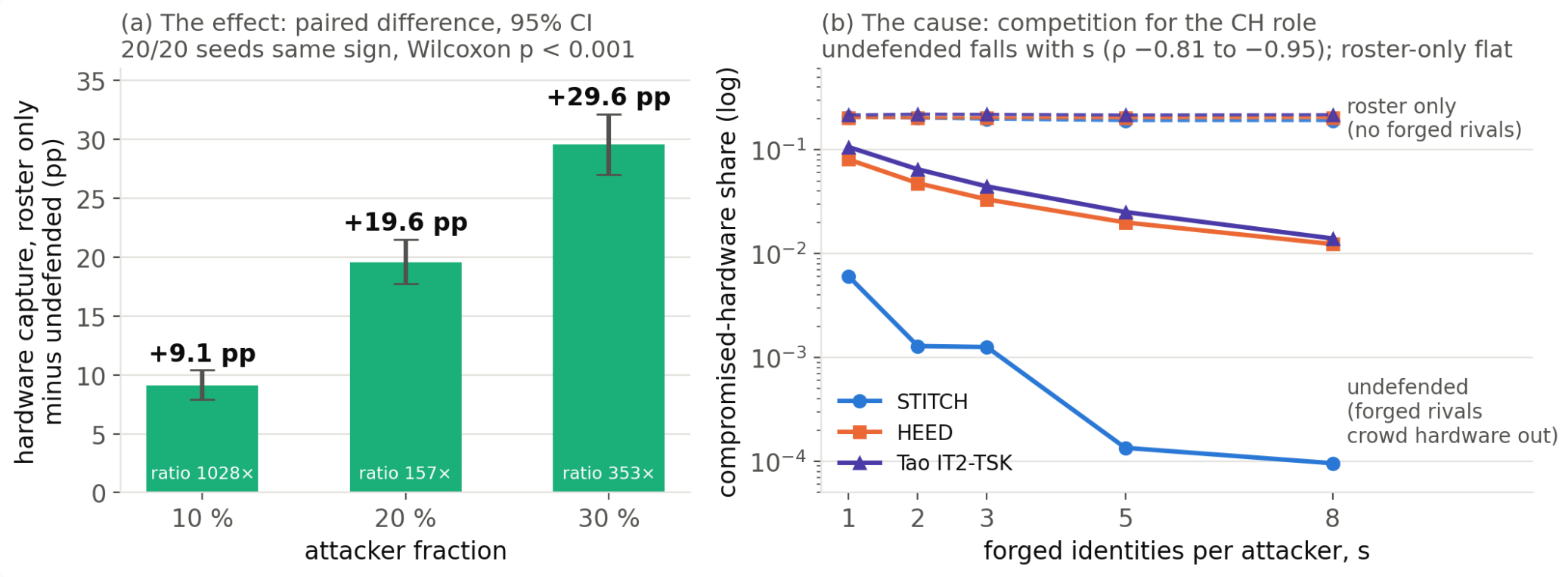
- **The roster is a complete Sybil defence and nothing else:** forged capture is zero in every round of every seed.
- **The veto is a complete hardware defence and no Sybil defence:** paired difference from undefended −0.0004 [−0.0012, +0.0004], p = 0.154.

20 % compromised	Forged-CH share	Hardware share	HND (rounds)
roster + veto	0.0000 [0.0000, 0.0000]	0.0007 [0.0006, 0.0009]	634.0 ± 45.6
veto only	0.1005 [0.0919, 0.1090]	0.0001 [0.0001, 0.0001]	533.9 ± 37.5
roster only	0.0000 [0.0000, 0.0000]	0.1971 [0.1791, 0.2163]	544.6 ± 44.5
neither	0.1010 [0.0922, 0.1096]	0.0013 [0.0002, 0.0028]	528.4 ± 38.4

20 seeds. Shares: mean [95 % bootstrap CI]. HND, half nodes dead: mean ± 95 % CI.

RESULTS · DISPLACEMENT

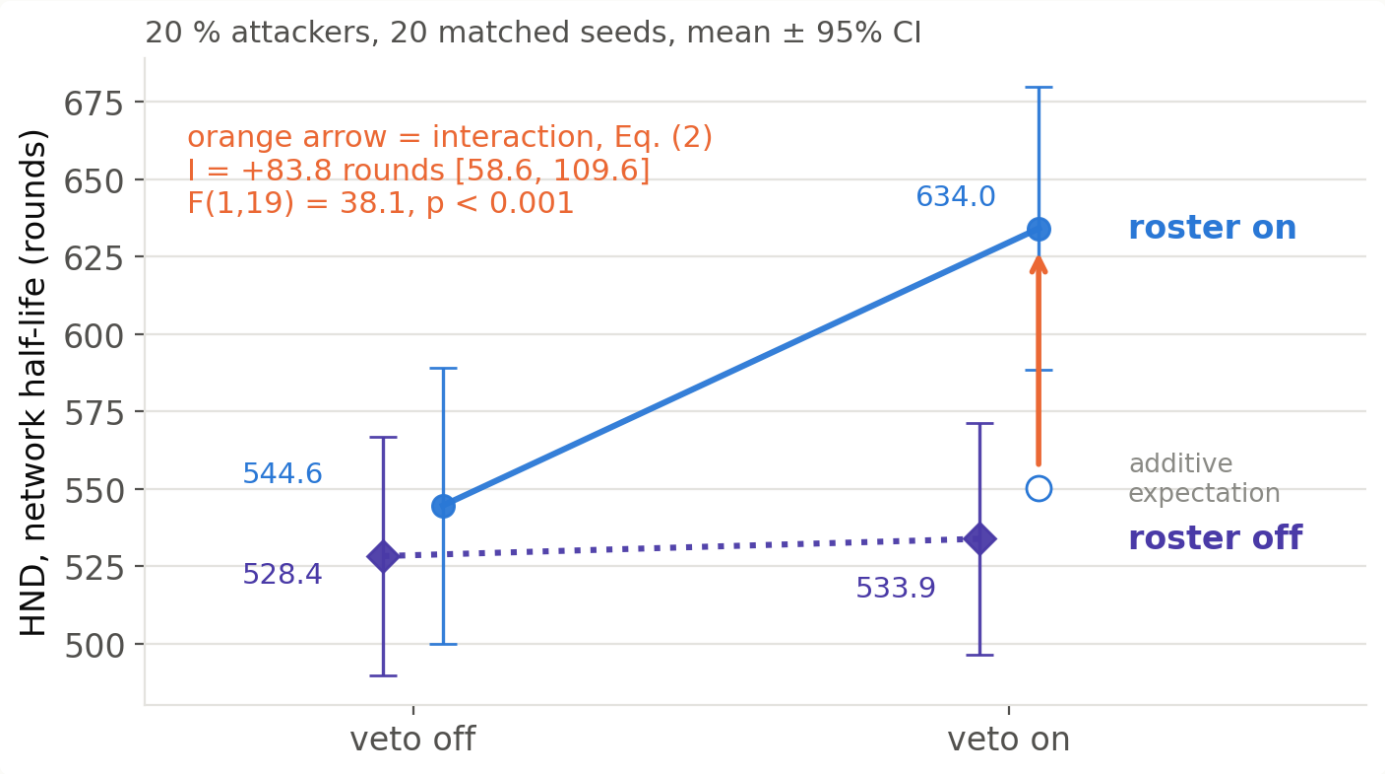
Partial coverage arms the uncovered channel



Roster alone: **+9 to +30 percentage points** of cluster-head roles to compromised hardware, all 20 seeds the same sign. The aggregate moves only 0.102 → 0.197.

RESULTS · NETWORK LIFETIME

Super-additive on network half-life



APART

veto **+1.0 %** · roster **+3.1 %**

TOGETHER

+20.0 %

Interaction, Eq. (2): I = HND(R+V) − HND(R) − HND(V) + HND(N). *A property of the host.*

ROBUSTNESS

The threat model is worth a factor of 190

Arm	Minting	Forged-CH share	Re-mints
veto only	one-shot	0.00053 ± 0.00005	0
veto only	churn	0.101 ± 0.009	4403 ± 294
roster + veto	one-shot	0.000 ± 0.000	0
roster + veto	churn	0.000 ± 0.000	0

20 % attackers, 20 seeds, mean ± 95 % CI.

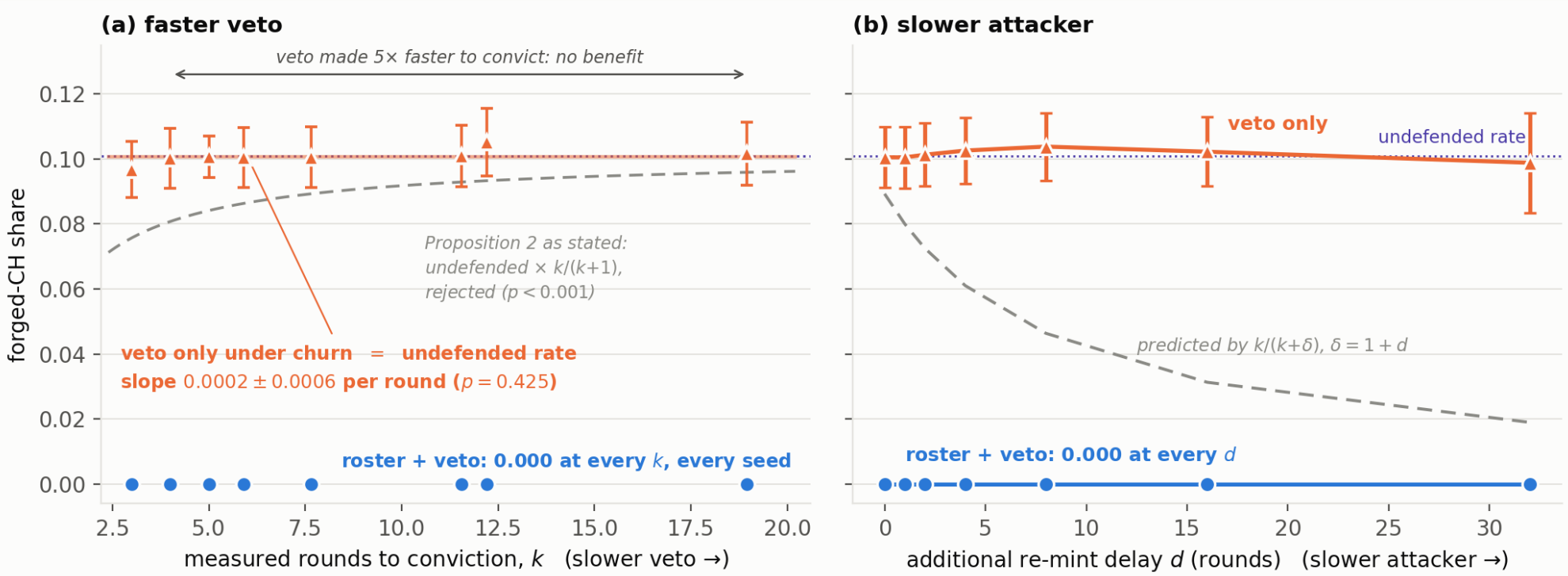
190×

worse when the attacker may re-mint, nothing else changed (all 20 seeds, Wilcoxon $p < 0.001$)

One-shot minting makes the veto *look* effective: an artefact of the threat model, not a property of the mechanism. The roster records **zero re-mints**: churn is futile against provenance.

ROBUSTNESS

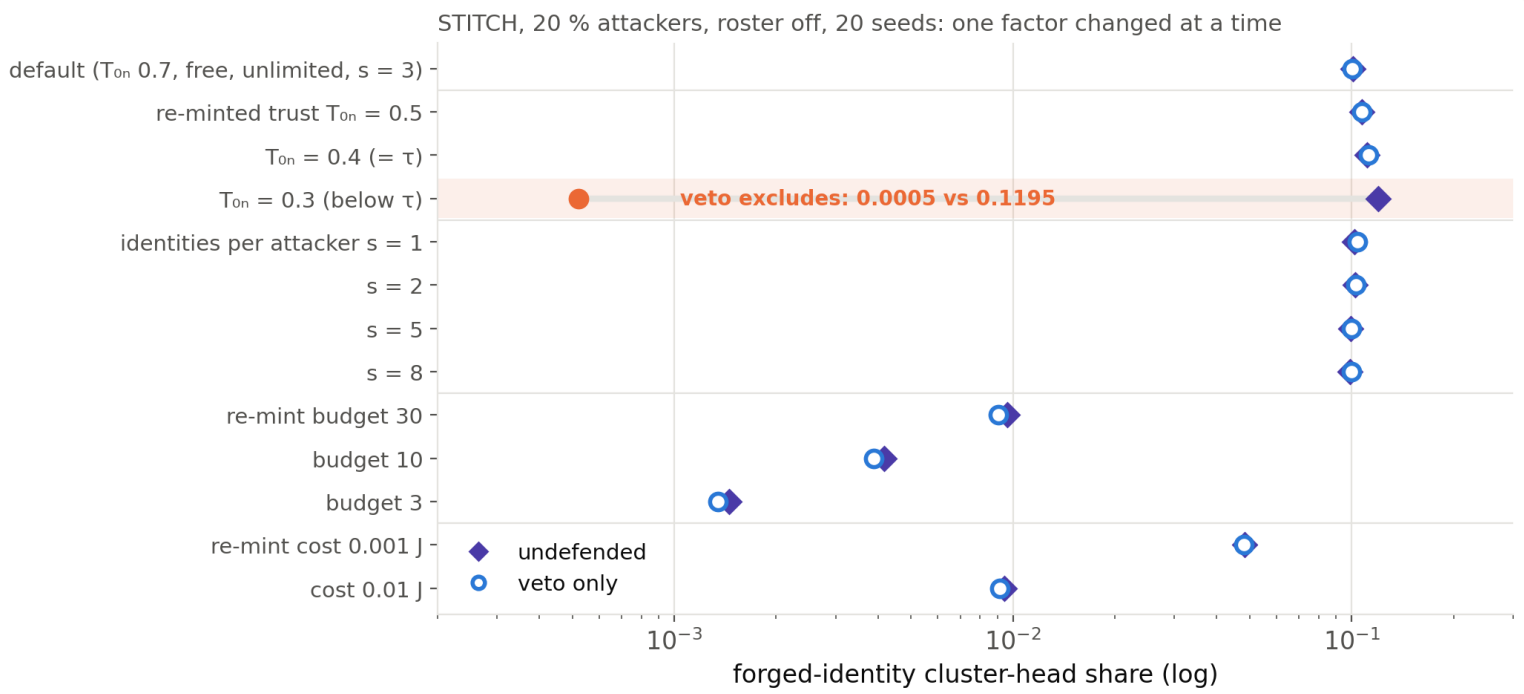
Neither a faster veto nor a slower attacker



(a) k from 18.95 to 3.00 rounds: slope 0.0002 ± 0.0006 (p = 0.425) against 0.0013 predicted (t = −3.35, p < 0.001). **(b)** d from 0 to 32 (δ 1 to 33): slope −0.00006 ± 0.00036 (p = 0.752); at d = 32, 0.0988 where k/(k+δ) predicts 0.0190.

ROBUSTNESS · SENSITIVITY ANALYSIS, 600 RUNS

What the flatness depends on



Not the pool: flat even at **s = 1** (0.1043 vs 0.1017).

Budget and cost cut capture in **both** arms alike.

Only a re-minted identity starting **below τ** lets the veto exclude. That is a newcomer quarantine: a **provenance** test, so P1 applies.

ROBUSTNESS · A CORRECTION WE FOUND OURSELVES

The quorum feeds the veto

CONSENSUS ROUND
validators vote; a failed-quorum proposer is penalised

→

EVIDENCE F, U
each validation counts for or against

→

TRUST T_i, EQ. (1)
$\eta = 0.2$, $\lambda = 0.5$, $T_0 = 0.7$

→

VETO GATE
may lead only if $T_i \geq \tau = 0.4$

CONSENSUS ON

0.0007

hardware capture · half-life 634.0 rounds

CONSENSUS OFF: THE VETO GOES BLIND

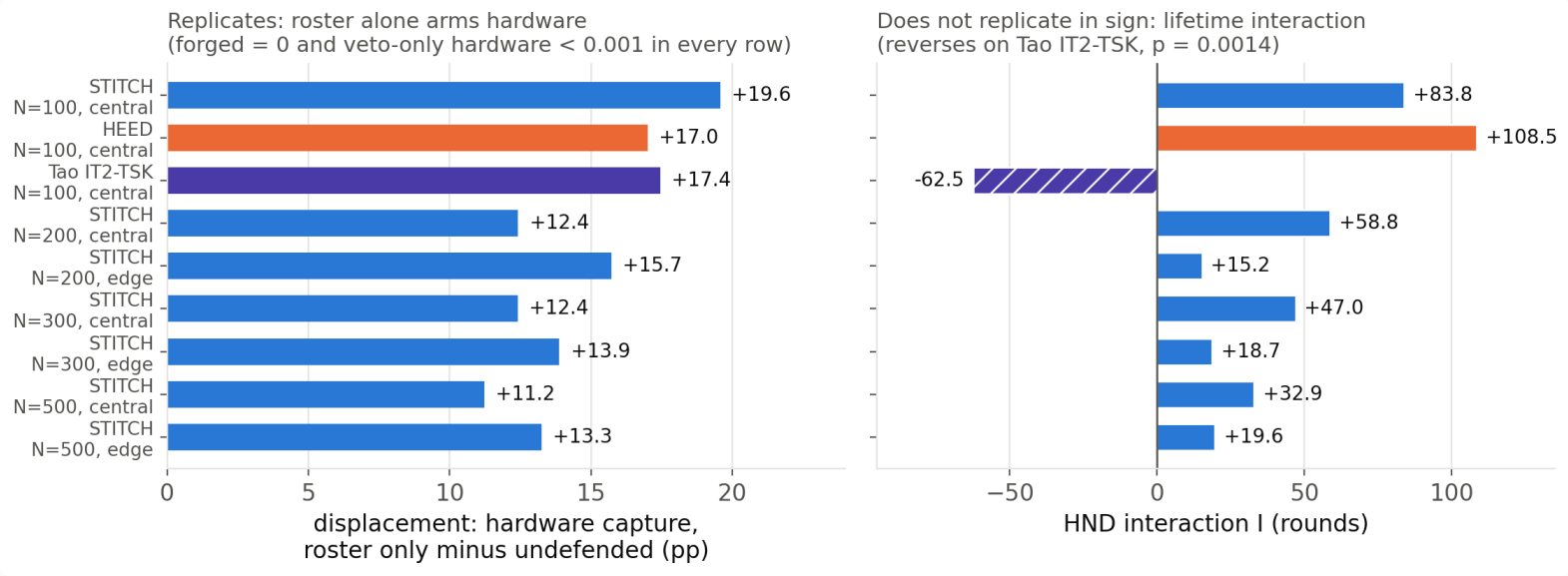
0.183

every trust drifts to $0.75 > \tau$ · half-life 561.5 rounds (-12.9 %, $p < 0.001$)

The quorum covers **neither population directly**, yet it is the veto's only evidence. *Attribution has to be measured, not assumed.*

ROBUSTNESS · SUPPLEMENTARY TABLE S1

Does it replicate?



REPLICATES, EVERY ROW

roster forged = 0; veto-only hardware < 0.001; roster alone arms hardware by 11 to 20 pp

DOES NOT

ratio size (4.9× to 157×); interaction sign; on Tao, veto-only forged 1.27× undefended

IMPLICATIONS

The attribution principle

1. **Provenance defends against renewable identities, and only those:** complete in every tested configuration.
2. **Behaviour defends against non-renewable entities, and only those:** while newcomer trust is at or above τ.
3. **A quorum covers neither directly, but may feed a mechanism that does:** worth 12.9 % of half-life here.
4. **Partial coverage can be worse than none on the uncovered channel:** 9 to 30 pp, on all three hosts.
5. **Attribution changes what a security claim means.**

> These mechanisms are complements, not alternatives.

IMPLICATIONS

A reporting standard

Each item is checkable by a reviewer. R1, R2 and R5 each overturn a conclusion in our data; R3 makes attribution possible at all.

#	Requirement
R1	Report forged-identity and compromised-hardware populations as separate metrics
R2	Grant the adversary identity churn by default; one-shot minting is the ablation
R3	Report a factorial ablation over the security mechanisms
R4	Report consensus or quorum activity separately from admission outcomes
R5	Where a defence or threat model has a time constant, report the outcome as a function of it

CONCLUSION

No aggregate metric can attribute a security result to a mechanism

ROSTER

0

forged capture in every tested setting; hardware left 19.7 ± 2.0 %

VETO

10.1 %

forged capture, the same as no defence (10.1 ± 1.0 %)

ROSTER ALONE

+9–30 pp

hardware capture: worse than no defence on that channel

BOTH

+20 %

half-life, super-additive; consensus costs ≈ 0.3 % of energy

When a paper says its protocol resists Sybil, ask: **which mechanism**, **against which population**, **what the attacker was not allowed to do**, and **what the uncovered population gained**.

Thank you

Roster or Reputation? Attributing Sybil Resistance to Mechanisms in Secure Clustered Wireless Sensor Networks: A Simulation Study

Saidi Adetunji Adegoke · K. A. Abdulsalam · N. A. Azeez · University of Lagos

3,960 runs · 20 seeds · 3 host protocols · code and data: doi 10.5281/zenodo.24244917

Questions?