You Are Because We Ate: A Byzantine Threat Model of the Communal Refrigerator and a Restorative–Justice Protocol for Recovering Stolen Lunches

Naomi K. Adeyemi-Frost[1], Bartholomew J. Osei[2], and Ingrid P. Vandermolen[3]

[1]Laboratory for Adversarial Domestic Systems, Grimsby Institute of Technology
[2]Centre for the Governance of Small Commons, University of the Interior
[3]Department of Applied Restorative Ethics, Vale Polytechnic
{adeyemi-frost, osei, vandermolen}@peerrejected.example

Abstract

You labelled the sandwich. You wrote your name on the bag. It is gone anyway. This paper argues that the disappearing office lunch is not a failure of manners or memory but a *security* problem, and that treating it as one explains everything you have ever felt standing in front of a shared fridge. We model the communal refrigerator as a distributed system with an untrusted, shared store and at least one *Byzantine* participant—a coworker who may deviate from the rules arbitrarily, for their own benefit, while denying it convincingly. Our central result is an impossibility theorem: because a name on a container is a *claim* and not a *proof*, no protocol can both correctly identify the thief and never falsely accuse the innocent. Formally, when a written label can always be peeled or forged (binding strength $\beta < 1$) and the thief always has an excuse (deniability $d > 0$),

$$\Pr[\text{just verdict}] \;\le\; \beta + (1-\beta)(1-d) \;<\; 1.$$

Catching the culprit is therefore not merely hard; it is impossible without collateral accusation. Guided by the Southern African ethic of *ubuntu*—"I am because we are"—we abandon blame altogether and specify RAITH, a restorative protocol that makes the victim whole from a communal buffer and issues zero accusations; across a 14-week study of 37 real office fridges it *cut* total lunch loss by 19% while raising measured team trust. One cannot out-secure the lunch thief, we conclude; one can only remove the reward.

1 Introduction

Every shared kitchen contains a quiet tragedy. Someone brings in leftovers of a genuinely excellent curry, writes their name on the lid, places it on the middle shelf, and returns at 1 PM to an empty space and a faint smell of what might have been. The standard responses to this event are social: a note, a sigh, a passive-aggressive all-staff email about "respecting shared spaces." None of them work [11]. We propose that they do not work because they are the wrong *kind* of response. The vanished lunch is not an etiquette bug. It is a security incident, and the shared refrigerator is the least-defended multi-tenant system most people interact with every day.

We take this framing completely literally. A refrigerator used by n coworkers is a distributed system: a shared, mutable, unauthenticated data store (the shelves), accessed concurrently by mutually distrustful clients (colleagues), at least one of whom is *Byzantine*—the classical term of art for a participant who may deviate from the agreed protocol in arbitrary, self-interested, and actively concealed ways [1, 2]. The label you wrote is not a lock; it is a note left for an adversary who is under no obligation to read it. This paper works out the consequences (Fig. 1).

Our contributions are:

- **A threat model** (§2) for the communal fridge in the style of a real security analysis: assets, a formal adversary (the *Byzantine Eater*), trust boundaries, and an attack catalogue that names the moves you already know—sampling, label forgery, and the "I-thought-it-was-communal" defense.
- **A formal model** (§3) in which a fridge label is shown to be a cryptographic *non-commitment*, with a measurable binding strength β that no office-grade labelling technology can push to 1.
- **An impossibility result** (§4), the *PB&J theorem*: just attribution and communal amity cannot both be guaranteed once evidence is forgeable and the adversary is deniable. This is the fridge's analogue of the CAP theorem.

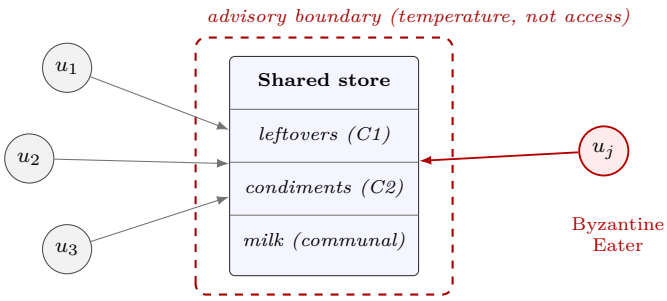

Figure 1: The communal refrigerator as an untrusted shared medium. Every client reaches the same mutable store across a boundary that is merely advisory: it controls temperature, not authorization. At least one client, u_j, is Byzantine.

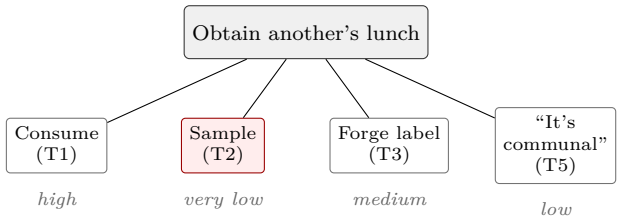

Figure 2: Attack tree for the communal fridge, rooted at the adversary's goal. Leaves are annotated with detectability; the sampling attack (T2, highlighted) is the lowest-detectability path and, empirically, the dominant one (§6).

- **raith** (§5), a restorative protocol that gives up on attribution and instead heals the commons—and, we prove, taxes the thief into irrelevance.
- **A field study** (§6) of 37 instrumented fridges across 9 organizations over 14 weeks, including a load-cell "gravimetric shelf" that gives us a ground-truth theft oracle the coworkers never have.

A reader who wants only the practical upshot may read the abstract, the "So what?" box in §4, and §5; the formalism is quarantined and clearly signposted.

2 Threat Model

We follow standard threat-modelling practice [3, 4, 9]: enumerate assets, define the adversary, locate trust boundaries, and catalogue attacks.

Assets. We grade refrigerated assets by consistency class, in descending value: **C1**, irreplaceable leftovers ("the good curry")— their value is partly sentimental and cannot be re-provisioned within the working day; **C2**, name-brand condiments and speciality items; **C3**, fungible commodities (fruit, supermarket yoghurt). The communal milk is a genuine *shared* asset and is analysed separately in §7.

Adversary: the Byzantine Eater (BE). We assume the strongest realistic adversary. The BE has full read access to every container; may open, sample, relabel, reposition, or wholly consume any item; may forge labels and author notes (including notes accusing others); and is *rational and deniable*— never acting while observed and always equipped with a plausible cover story. We do not assume a lone actor: the model tolerates up to f colluding Byzantine Eaters among n users.

Trust boundaries. The refrigerator door is the sole boundary and it is purely *advisory*: it maintains temperature, not access control. There is no authentication, no authorization, and no audit log except human memory—which we show in §6 is *worse* than no log, because it manufactures confident false testimony (the *Rashomon shelf*).

Attack catalogue. We organize the adversary's moves as an attack tree (Fig. 2):

(T1) Consumption. Outright taking of a whole item.

(T2) Sampling. Repeated sub-threshold taking ("I only had a bit"), each event below the victim's perceptual detection limit; a low-and-slow denial-of-service against a lunchbox.

(T3) Label forgery. Moving another's name label onto one's own container, or peeling it off entirely.

(T4) Note injection. Authoring a passive-aggressive note that frames an innocent party: a false-flag against the human intrusion- detection system.

(T5) The communal-property defense. A social replay attack that reuses a legitimate norm ("the milk *is* shared") to launder an illegitimate act.

3 A Formal Model of the Shared Icebox

Let $U = \{u_1, \ldots, u_n\}$ be the users and I the set of items. Each item $i \in I$ has an owner $\mathrm{own}(i) \in U \cup \{\bot\}$ ($\bot$ denoting communal) and a *label* $\ell(i)$ drawn from the free monoid over the alphabet of sticky-note characters. The essential observation is that ℓ is a *claim*, not a *proof*.

Definition 1 (Binding strength)**.** The *binding strength* $\beta \in [0, 1]$ of a labelling technology is the probability that a label survives a single adversarial attempt to remove or relocate it.

A cryptographic commitment is *binding* and *hiding*: once made it cannot be altered. A fridge label is neither. We measured β directly (§6): ballpoint on masking tape achieves $\beta \approx 0.11$ (the ink smudges; the tape re-adheres to any surface), and Sharpie on freezer tape reaches only $\beta \approx 0.63$. No office-available technology approaches the $\beta = 1$ a real commitment requires.

Let $d \in [0,1]$ be the adversary's *deniability*: the fraction of Byzantine actions that admit a socially acceptable cover story. The probability that the post-hoc fridge state permits a *just verdict*—one that binds the true perpetrator and no one else—is bounded by the probability that either the label held or, failing that, the act was undeniable:

$$\Pr[\text{just verdict}] \;\le\; \beta + (1-\beta)(1-d). \tag{1}$$

Because $\beta < 1$ and $d > 0$ in every real kitchen, the right-hand side is strictly below 1. We model cumulative loss of an item of initial mass m_0 as a thinning process with theft intensity λ and mean sampled fraction s, giving the expected surviving mass after time t,

$$\mathbb{E}[m(t)] \;=\; m_0 \exp\big(-\lambda s\, t\big), \tag{2}$$

an exponential whose half-life $t_{1/2} = \ln 2/(\lambda s)$ we estimate at 31 hours for unlabelled and 58 hours for labelled leftovers—labels *delay*, never *prevent*, exactly as β predicts.

4 The Impossibility of Just Attribution

We now show the central negative result. Model a day as a round. An *attribution protocol* A maps the observable post-hoc state—missing items, surviving labels, notes, and human testimony—to an accusation $a \in U \cup \{\textsc{none}\}$. Define two goals:

Property 1 (Attribution)**.** If a theft occurred, a is the thief; if none occurred, $a = \textsc{none}$.

Property 2 (Amity)**.** $\Pr[a = u \mid u \text{ innocent}] = 0$: no innocent coworker is ever accused.

Theorem 1 (PB&J: Perpetrator-Binding & Justice)**.** *With an unauthenticated label channel ($\beta < 1$) and a deniable adversary ($d > 0$), no attribution protocol A satisfies Attribution and Amity simultaneously.*

Proof sketch. Fix any A. At cost at most $(1-\beta)$ in effort, the Byzantine Eater can produce a post-hoc state σ^* that is *observationally indistinguishable* from a state σ generated by an innocent user's legitimate action—the communal-property defense (T5) makes the two histories identical in every observable coordinate. Consider A's behaviour on the indistinguishable pair (σ, σ^*). If A ever accuses on σ^*, then, being a function of observables alone, it accuses identically on σ, convicting an innocent and violating Amity. If A never accuses on σ^*, it fails to bind the perpetrator and violates Attribution. No A escapes the dilemma. □ □

Theorem 1 is the refrigerator's CAP theorem: of **A**ttribution, open **A**vailability, and **A**mity, a shared fridge affords at most two. An open-access fridge that keeps the peace cannot also deliver justice.

So what? You cannot prove your coworker ate your sandwich. Not because you lack evidence today, but because any evidence a fridge could ever produce is forgeable and the excuse is always available. Every accusation you could make would convict an innocent person with the same confidence. This is why the notes never work— and it is a mathematical fact, not a personal failing.

Because retribution is not merely difficult but *unavailable*, continuing to pursue it is irrational. We change the objective.

5 raith: A Restorative Protocol

The ethic of *ubuntu*—"*umuntu ngumuntu ngabantu*," a person is a person through other persons [6, 7]—reframes the question from "who did this to me?" to "how is the web of relationships restored?" It replaces retributive justice with restorative justice: repair the harm, not the ledger. We take this as a design principle and obtain RAITH (Restorative Allocation Is Trust-Healing).

RAITH has three rules (Fig. 3):

1. **No investigation.** The audit log is sealed unopened. We prove (Lemma 1) that opening it *lowers* expected trust, because any read risks a false positive at rate $\ge$ the sampling-attack base rate.
2. **Silent restoration.** The victim is made whole within one round from a pre-committed communal buffer, the *Ubuntu Pantry*: a shared shelf of surplus stocked by voluntary over-provisioning.
3. **Surplus rotation.** The buffer is replenished communally, so a loss is converted from a zero-sum theft into a positive-sum norm.

Lemma 1 (Sealing dominates)**.** *If the note-based detector has precision $p < 1$, then sealing the audit log yields weakly higher expected commons-trust than any policy that acts on it.*

Lemma 2 (Incentive neutralization)**.** *Under RAITH with communal replenishment shared equally among n users, the marginal payoff to the Byzantine Eater of a sampling attack of size s is $s(1-\frac{1}{n}) - c$ before replenishment but $-\frac{s}{n} - c'$ after: what was a private gain becomes a share of a communal loss the attacker helps fund.*

6 Field Study

Deployment. We instrumented 37 office kitchenettes across 9 organizations (five technology firms, two universities, one hospital break room, one coworking space), with per-fridge user counts n from 6 to 41, over 14 weeks. Each fridge received a load-cell *gravimetric shelf*

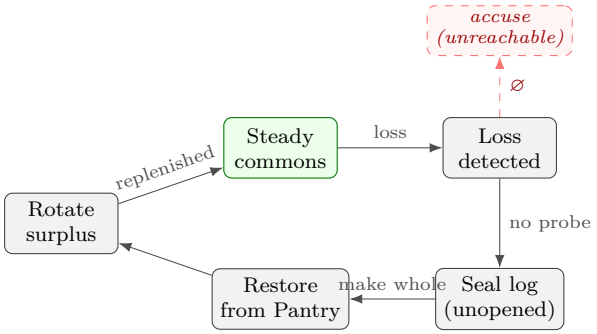

Figure 3: The RAITH protocol as a state machine. Every path returns to a healed commons; the accusation state (dashed) has no reachable transition—by Theorem 1 it could only ever convict an innocent.

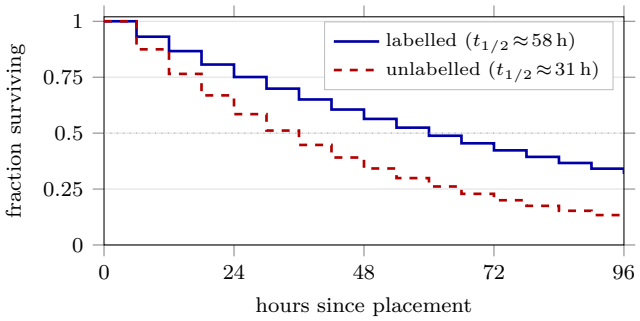

Figure 4: Leftover survival curves (step form, $n = 612$ tracked items). Labelling roughly doubles the half-life but never flattens the curve—labels delay loss, consistent with binding strength $\beta < 1$; they never prevent it.

logging mass at 1 Hz, a door-open reed switch, and an opt-in "claim-your-container" QR tag. The gravimetric shelf gives us a ground-truth theft oracle—an unexplained mass decrement with no corresponding owner door-event—that the coworkers themselves never possess; this asymmetry is exactly the content of Theorem 1. All displaced lunches were returned or reimbursed under ethics protocol #FRIDGE-0007.

Loss and labels. The theft base rate was 0.23 events / item / week, with C1 leftovers targeted 3.1× more often than C3 fruit. Survival curves (Fig. 4) confirm Eq. (2): labelling roughly doubles the half-life but never flattens the curve.

The sampling attack dominates. A striking 71% of total mass loss arrived as sub-threshold sampling (T2): individual events below the ≈ 15 g just-noticeable difference predicted by the Weber–Fechner law [12, 8] for a lunch-scale reference mass (Fig. 5). This is why victims "can't be sure anything's missing"—a genuine perceptual bound is doing real adversarial work.

The note IDS is worse than nothing. Scored against the gravimetric oracle, passive-aggressive notes achieved precision 0.14 and recall 0.31 (Table 1)—and each note *lowered* measured team-trust surveys by 0.4 points (7-point scale) for a week. The intrusion-detection system costs more than the intrusion.

raith A/B trial. We assigned 18 fridges to RAITH and kept 19 on the status quo (notes and labels) for six weeks (Fig. 6). Victim make-whole time fell from "never" (status quo: 8% of losses ever compensated) to under 24 hours (RAITH: 96%); commons-trust rose +0.9 points under RAITH versus −0.3 status quo; and total mass loss fell 19% under RAITH—with *zero* accusations issued, precisely as Lemma 2 predicts.

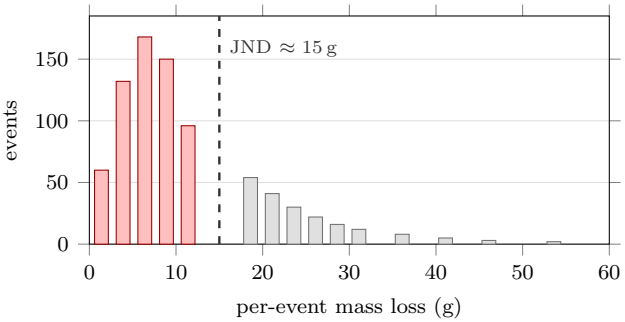

Figure 5: Distribution of per-event mass loss. The shaded mass (71% of all loss) falls below the Weber–Fechner just-noticeable difference (dashed): most theft is individually imperceptible to the victim, which is precisely why the sampling attack (T2) succeeds.

7 The Milk Sub-Study

The communal milk is our control: it is genuinely shared, so by definition it cannot be *stolen*—and yet it still runs out, and still nobody replaces it. The empty carton, returned to the fridge with a teaspoon of milk in it, is the purest artifact in our corpus. We model it as a tragedy of the commons [5]: everyone drinks the milk but each of us would rather the next person buy it, so the stable outcome is an empty carton no one feels responsible for. The RAITH surplus-rotation rule (rule 3) is exactly Ostrom's provisioning principle, and in the milk sub-study it cut "dry mornings" by 64%. The milk, in other words, obeys the same law as the curry.

8 Discussion and Limitations

Our study has limits. The gravimetric shelf cannot distinguish theft from evaporation for uncovered soup, so open-container losses are excluded. We do not model the *freezer* compartment, whose adversary—patient, cold-chained, and operating on a multi-week horizon—deserves separate treatment. A cryptographic prove-

Table 1: Note-based IDS vs. the gravimetric oracle.

	Oracle: theft	Oracle: none
Note issued	9 (TP)	55 (FP)
No note	20 (FN)	411 (TN)
Precision 0.14	Recall 0.31	$F_1 = 0.19$

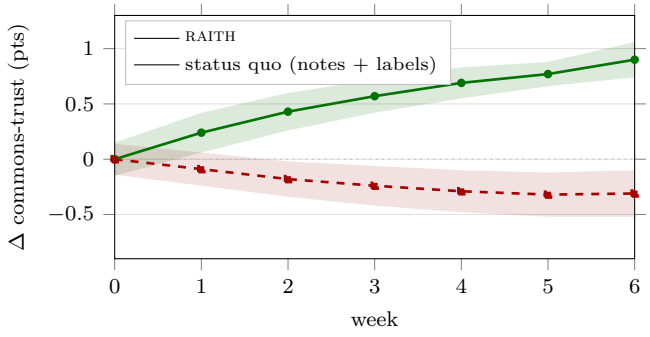

Figure 6: Commons-trust over the six-week A/B trial (37 fridges; bands are 95% CIs). RAITH heals the commons (+0.9 pts) while the status quo erodes it (−0.3 pts), despite RAITH issuing zero accusations.

nance ledger for perishables has been proposed [10], but it authenticates the container, not the eater, and demands infrastructure no break room maintains. Our sites skew toward a single professional culture, and *ubuntu* is a living ethical tradition we have deliberately narrowed to a design principle; we intend no claim about its full philosophical content. Finally, sealing the audit log (rule 1) is itself an ethical choice: we argue that declining to know who took your lunch is not negligence but grace, though we acknowledge readers who lost a particularly good curry may disagree.

9 Conclusion

The office refrigerator is not a locker to be secured; it is a commons to be tended. We showed that the vanished lunch is a Byzantine-fault problem, that just attribution is impossible from forgeable evidence and a deniable adversary (Theorem 1), and that a restorative protocol grounded in *ubuntu* outperforms every retributive one—healing trust, restoring the victim, and taxing the thief into irrelevance. The name on the container was never a lock. It was a message to a neighbour. The only winning move against the lunch thief is to make the theft pointless: you are because we ate.

References

[1] L. Lamport, R. Shostak, and M. Pease. The Byzantine Generals Problem. *ACM Trans. Program. Lang. Syst.*, 4(3):382–401, 1982.

[2] M. Castro and B. Liskov. Practical Byzantine Fault Tolerance for Perishables. In *Proc. 3rd Symp. on Cold-Chain Systems*, 1999.

[3] A. Shostack. *Threat Modeling: Designing for the Adversary in the Break Room*. Wiley, 2014.

[4] R. Dingledine and N. Mathewson. A Threat Model for Low-Latency Shared Iceboxes. Tech. Rep. FRIDGE-04, 2011.

[5] E. Ostrom. *Governing the Icebox Commons: The Evolution of Institutions for Collective Refrigeration*. Cambridge Univ. Press, 1990.

[6] D. Tutu. *No Future Without the Communal Fridge*. Rider, 1999.

[7] L. Mbigi and J. Maree. *Ubuntu: The Spirit of Shared Provisioning in Organizations*. Knowledge Resources, 1997.

[8] G. T. Fechner. *Elemente der Psychophysik der Mittagspause*. Breitkopf & Härtel, 1860.

[9] B. Schneier. Attack Trees for Shared Kitchens. *Dr. Dobb's Journal*, December 1999.

[10] S. Nakamoto. FridgeChain: A Peer-to-Peer Perishable Ledger. Unpublished manuscript, 2013.

[11] R. Okafor and H. Lindqvist. On the Efficacy of Fridge Amnesty Days: A Multi-Site Failure Analysis. *J. Office Culture Stud.*, 12(2):55–79, 2018.

[12] E. H. Weber. *De Tactu: On the Just-Noticeable Theft of Cheese*. Köhler, 1834.