When One Leak Pays Forever: Context Binding and the Price of Deterring Collusion

Tingyi Lin[1†], **Shawn Yu**[2], **Ruoran Lai**[3], **Huanxi Zhang**[4]
[1]Tropic Research Institute [2]Boston College
[3]Sun Yat-sen University [4]University of Wisconsin–Madison

Abstract

A coalition that deviates once can profit many times when what it sells keeps working. In a threshold-encrypted mempool, a leading defense against maximal extractable value (MEV), a quorum of the decryption committee that sells its decryption capability to a front-runner exposes every later block that the capability still decrypts. We ask how large a penalty, such as slashable stake, deters this kind of collusion. In our repeated game, a single leak by any coalition in a monotone family of authorized coalitions (for example, any k of the n committee members) unlocks a set of future rounds, costs a one-time penalty, and ends the coalition's participation. We show that every dynamic deviation reduces to choosing a leak time, so deterrence holds if and only if each coalition's penalty covers the largest discounted value that a single leak reaches. Without discounting, over T rounds of unit value full reuse needs a penalty of T while binding each leak to its own round needs 1, so no penalty that is constant in the horizon deters unbounded reuse; a reuse window of w rounds costs at most w times the largest per-round value. The cheapest profile of per-party stakes that deters every coalition solves a covering linear program. For blockchain design, per-epoch keys cut the required stake from the value of a key's lifetime to the value of one epoch; we calibrate the gap on Ethereum front-running data and place Ferveo and Shutter in the model. The analysis extends to sealed-bid auctions, multi-authority voting, and federated learning under a shared key.

1 Introduction

Three may keep a secret, if two of them are dead.
— Benjamin Franklin, *Poor Richard's Almanack*, 1735

Deterrence is usually priced one act at a time: a deviation is deterred when the penalty it risks exceeds what it gains [6]. In a repeated game the gain is a continuation value [22], and the comparison stays simple as long as each deviation is a single act. It stops being simple when a coalition deviates once by selling a capability that keeps working. The buyer then collects in later rounds as well, the coalition's gain becomes a stream, and the penalty needed to deter the sale depends on how many future rounds that one sale reaches.

Threshold-encrypted mempools are the leading example [5, 16, 19]. Users encrypt transactions to a committee key, and a batch is decrypted only after its order is fixed, which withholds the information that front-running and sandwich attacks exploit [14, 33]. Any authorized quorum of the committee can still reconstruct a decryption capability early and sell it to a front-runner. When the committee key lives for many epochs, one sale reveals every later batch encrypted under it. Context-dependent threshold decryption [8] binds each decryption share to a public context such as an epoch number.

[†]Correspondence to: tingyi3@illinois.edu

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

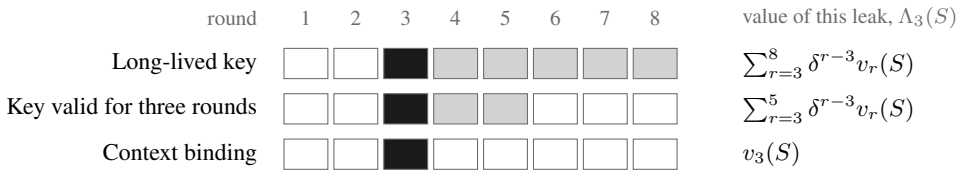


Figure 1: A leak at round 3 under three keying designs. Dark: the round of the leak; gray: later rounds whose data the buyer can still read; white: rounds that stay hidden. The right column is the value $\Lambda_3(S)$ of this leak, and Theorem 4.1 requires $\kappa(S) \geq \sup_t \Lambda_t(S)$.

When every epoch has its own key, or when the long-term key shares sit in modules that decrypt only the current context's ciphertexts and are replaced as soon as a leak becomes public, a sale in one epoch reveals nothing about the next. Sealed-bid auctions, multi-authority voting, and federated learning under a collective key share this pattern of one leak and many rounds (Section 6).

We study a repeated game among n parties who jointly hold a decryption capability. A monotone family $\mathcal{A}$ of *authorized* coalitions, such as all coalitions of at least k parties, lists who can leak. The *stage value* $v_t(S)$ is what coalition S can earn by selling early access to the hidden data of round t. A leak is public and ends the leaking parties' role in the protocol, so each coalition leaks at most once. The *reachability set* $\mathcal{R}_S(t)$ is the set of rounds whose data a buyer can read after S leaks at round t; a long-lived key makes it every later round, and context binding, which ties everything a coalition can leak to its round, makes it round t alone (Figure 1). The *exposure* $\kappa(S)$ is the one-time penalty that S bears for its leak, such as slashed stake, a tracing liability, or forfeited rewards. Values are discounted by a factor δ per round and are transferable within a coalition, and the discounted value that a single leak reaches is bounded across leak times. A coalition is *deterred* if, after any history without a leak, it gains nothing from any deviation strategy, and honest execution is *coalition-safe* if every coalition is deterred.

Our first result prices deterrence exactly.

Theorem A (Exact deterrence threshold)**.** *A coalition S is deterred if and only if its exposure $\kappa(S)$ is at least the discounted value $\Lambda_t(S)$ of the rounds that a leak by S at round t reaches, for every t:*

$$\kappa(S) \geq \sup_t \Lambda_t(S), \qquad \Lambda_t(S) = \sum_{r \in \mathcal{R}_S(t)} \delta^{r-t} v_r(S).$$

Honest execution is coalition-safe if and only if this holds for every authorized coalition.

Theorem A is Theorem 4.1. A coalition may time its leak to any history of play, and one leak may unlock a nonconsecutive set of later rounds. Because the leak is the coalition's last act, every deviation is worth exactly what its leak is worth (Lemma A.1), so the repeated interaction enters only through the reachability sets. The one-shot rule of Becker [6] is the case $\mathcal{R}_S(t) = \{t\}$.

Our second result measures what reuse costs.

Theorem B (Reuse sets the price)**.** *Fix an authorized coalition S.*

(i) On a T-round instance without discounting in which every stage value of S equals 1, deterring S requires exposure T when a leaked capability stays valid in every later round, and exposure 1 when it is bound to its own round.

(ii) Without discounting and with unit stage values, S is deterred exactly when its exposure is at least $\rho(S)$, the largest number of rounds that one leak by S reaches. No exposure that is constant in the horizon therefore deters S along a family of instances in which $\rho(S)$ grows without bound.

(iii) If every leak by S reaches at most $\rho(S)$ rounds and every stage value of S is at most $\bar{v}(S)$, then exposure $\bar{v}(S) \sum_{j=0}^{\rho(S)-1} \delta^j$, which is at most $\rho(S)\, \bar{v}(S)$, deters S. No smaller exposure does when a leak at some round t reaches exactly the rounds $t, \ldots, t + \rho(S) - 1$, each of value $\bar{v}(S)$.

Our third result splits the requirement across parties.

Theorem C (Cheapest deterring stakes)**.** *Suppose party i posts exposure $d_i \geq 0$, for example bonded stake, and a coalition's exposure is the sum over its members. A profile (d_i) deters every authorized*

coalition exactly when it satisfies one covering constraint per authorized coalition, so the cheapest such profile solves a covering linear program. If the authorized coalitions are those with at least k members, for some $1 \le k \le n$, all parties post the same amount, and the supremum of a coalition's reachable value depends only on its size m, say λ_m, then the smallest deterring amount per party is $\max_{k \le m \le n} \lambda_m / m$.

Theorem B collects Theorem 4.5, Proposition 4.6, and Corollary 4.7; Theorem C collects Corollaries 4.8 and 4.9. Section 5 extends Theorem A to probabilistic and time-distributed exposure, where liability that grows with use restores deterrence under long reuse.

Techniques. Three objects describe a leak: who can produce it ($\mathcal{A}$), what each round is worth (v_t), and how far one leak propagates ($\mathcal{R}_S$). Only the third involves time. Lemma A.1 shows that it carries all of the dynamics, and exact deterrence becomes a supremum over first-leak times. Two smaller state variables fail. Comparing $\kappa(S)$ with the largest stage value, as a one-shot model does, is exact under context binding and can be off by a factor of T under full reuse (Theorem B(i)). Counting reachable rounds is exact with unit values and no discounting (Proposition 4.6); otherwise the identity and timing of the reached rounds matter. Since the reachability set is the only dynamic object, context binding acquires a one-line game-theoretic meaning, $\mathcal{R}_S(t) = \{t\}$, and Theorem B prices it.

2 Related Work

Deterrence and coalition deviations. Comparing an expected penalty with the gain from an offense goes back to Becker [6], and Corollary 4.2 is that rule for a single leak. Deviations by groups are the subject of strong equilibrium [3] and coalition-proof equilibrium [7]. Our benchmark is lighter: a coalition deviates from honest execution and, since utility is transferable inside it, acts as one player. In repeated games the one-shot deviation principle reduces deviations to single-period ones when payoffs are continuous at infinity [22]. Lemma A.1 plays that role here, with one difference: a leak is irreversible, so the deviation that matters is the first leak, and the reachability set fixes its value. The new element is this state variable: it suffices for exact deterrence over any monotone family of coalitions, and it separates reusable from context-bound deviations by a factor of T.

Stake, slashing, and economic security. Proof-of-stake protocols deter misbehavior by slashing bonded stake [11], and Budish, Lewis-Pye, and Roughgarden [10] study when attacks on permissionless consensus can be made expensive. In those analyses the slashed stake is weighed against the profit of an attack. Our exposure κ plays the role of the slashed stake, and the profit it must outweigh is a stream over every round that one leak unlocks.

Incentives in collaborative learning. Mechanisms for collaborative and federated learning deter individual participants who under-collect data, fabricate it, or send dishonest updates [15, 12]. The deviation studied here is collective, a quorum that can open data encrypted under a shared key, and its payoff accrues over every round the key protects.

Rational cryptography. Rational secret sharing [18, 1] and rational cryptography more broadly [24] ask when cryptographic protocols implement desired equilibria among self-interested parties. There the parties reconstruct or compute a secret among themselves. Here an authorized coalition sells a functional decoder to an outsider, and the difficulty is the continuation value of a decoder reused across contexts, which one-shot reconstruction never faces.

Threshold decryption, accountability, and context binding. Ferveo [5], Shutter [16], and BlindPerm [19] hide transaction content until its order is fixed, and ETHTID [31] combines scheduled disclosure with stake-backed penalties. Threshold traitor tracing [9] traces a leaked decoder to a member of the quorum that produced it, and context-dependent threshold decryption [8] ensures that decryption shares issued under different contexts cannot be combined, a guarantee against fewer than a threshold of corrupt parties. These works build and analyze the cryptography. In our model their accountability mechanisms determine the exposure κ, and their keying choices, together with what a quorum can hand to a buyer, determine the reachability sets (Section 6.2).

MEV and order fairness. Transaction reordering and front-running were identified by Daian et al. [14] and measured at scale by Torres et al. [33]; Kulkarni et al. [20] study their game-theoretic

structure in constant-function market makers. Order fairness, inclusion lists, and protected order flow are analyzed in [21, 34, 4], the economics of mempool privacy in [28], and mitigation more broadly in the surveys [2, 23]. Accountable mempools such as LØ [26] act on the exposure κ or on its time-distributed version. We take the stakes of these applications as the stage values $v_t(S)$ and study the leak that precedes honest release.

3 Model and Preliminaries

3.1 Access Structures and Repeated Contexts

We fix a set of parties $[n] = \{1, \dots, n\}$.

Definition 3.1 (Monotone access structure)**.** A family $\mathcal{A} \subseteq 2^{[n]} \setminus \{\varnothing\}$ is a *monotone access structure* if $S \in \mathcal{A}$ and $S \subseteq S'$ imply $S' \in \mathcal{A}$. A coalition S is *authorized* if $S \in \mathcal{A}$ and *unauthorized* otherwise.

In the language of cooperative game theory, $\mathcal{A}$ is the family of winning coalitions of a monotone simple game [30, 32]. A k-out-of-n policy is the weighted voting game with unit weights and quota k, and a committee whose shares are weighted by stake is a general weighted voting game.

The parties run a service over rounds $t = 1, 2, \dots, T$, with horizon $T \in \mathbb{N} \cup \{\infty\}$ and discount factor $\delta \in (0, 1]$ per round. Round t has a public context tag χ_t, whatever a leaked capability can be bound to (an epoch or batch in an encrypted mempool, an auction instance, a training round), and hidden data that are due to become public at that round's honest release point.

3.2 Leakage Values and One-Time Accountability Exposure

For each coalition $S \subseteq [n]$ and round t, let $v_t(S) \in \mathbb{R}_{\geq 0}$ denote the coalition's *stage value*: its transferable-utility gain from an outside buyer's unauthorized early access to round t's hidden data, whether a leak by S at round t or an earlier one provides that access. It is a summary statistic that may absorb side payments from a buyer, front-running rents, or any other value of early access, and the theorems need no decomposition of it.

Accountability is captured by a public *one-time exposure function* $\kappa : 2^{[n]} \to \mathbb{R}_{\geq 0}$: a coalition S that leaks bears the loss $\kappa(S)$, measured in present value at the round of the leak. This loss may represent slashable stake, legal penalties, lost future rewards, reputational damage, or an effective expected penalty after probabilistic tracing.

Remark 3.2. The exposure is an effective quantity. If tracing identifies at least one member of S with probability $p(S)$ and the present-value fine after a successful trace is $f(S)$, the risk-neutral exposure is $\kappa(S) = p(S)f(S)$, the coalition's total present-value liability from the leak.

3.3 Reachability Semantics

The central modeling object is the set of future contexts compromised by one leak. A *successful leak*, the sale of a functional decoder to a buyer, is public, and we assume that it is terminal: the leaking parties take no further part in the protocol, as when they are ejected from the committee. Each coalition therefore leaks at most once.

Definition 3.3 (Leakage reachability)**.** For each authorized coalition S and round t, let $\mathcal{R}_S(t)$, a subset of $\{t, t+1, \dots, T\}$ that contains t, be the set of rounds whose hidden data become accessible to S's buyer as a result of a successful leak by S in round t. For unauthorized coalitions we set $\mathcal{R}_S(t) = \varnothing$.

Since the leak is terminal, $\mathcal{R}_S(t)$ records everything it yields: every later round when the leaked object stays valid forever, and round t alone when nothing the coalition can leak opens another context.

Definition 3.4 (Discounted leakage value)**.** For every coalition S and round t, define

$$\Lambda_t(S) = \sum_{r \in \mathcal{R}_S(t)} \delta^{r-t} v_r(S) \in [0, \infty].$$

Throughout, we assume that $\sup_{t\le T}\Lambda_t(S)<\infty$ for every $S\in\mathcal{A}$. The assumption holds automatically when $T<\infty$, and when $T=\infty$ with $\delta<1$ and bounded stage values; when $T=\infty$ and $\delta=1$, it requires the undiscounted sums $\sum_{r\in\mathcal{R}_S(t)} v_r(S)$ to be bounded in t.

Thus $\Lambda_t(S)$ is the present-value benefit, at round t, of a leak at round t. When the supremum is infinite, no finite one-time exposure deters S.

Definition 3.5 (Reusable and context-isolating schemes)**.** We say that a deployment is:

(i) *fully reusable* if $\mathcal{R}_S(t)=\{t,t+1,\dots,T\}$ for every authorized coalition S and every round t;

(ii) *context-isolating* if $\mathcal{R}_S(t)=\{t\}$ for every authorized coalition S and every round t.

Remark 3.6. Context isolation, which we also call context binding, says that nothing an authorized coalition can leak in context χ_t is useful in any context $\chi_{t'}\neq\chi_t$. Context-dependent threshold decryption [8] ensures that decryption shares issued under different contexts cannot be combined, which protects a ciphertext against fewer than a threshold of corrupt parties. An authorized quorum, however, also holds long-lived key shares that decrypt under every context. The scheme is context-isolating when these shares sit in *revocable modules*, which decrypt only ciphertexts bound at encryption to the current context and which the protocol revokes and replaces before the next context once a leak becomes public, whoever supplied it; shares that merely cannot be exported do not suffice, because the modules themselves can be handed to a buyer.

3.4 Deviation Benchmark

The object of interest is whether honest execution is safe against coalition deviations under transferable utility.

Definition 3.7 (Deterrence and coalition safety)**.** Fix $(\mathcal{A},\{v_t\},\kappa,\{\mathcal{R}_S(t)\},\delta)$. A coalition $S\subseteq[n]$ is *deterred* if, in every round t and after every public history in which no successful leak has occurred before round t, no deviation strategy gives S a positive gain in transferable utility, in present value at round t, over continuing to behave honestly forever. Honest execution is *coalition-safe* if every coalition is deterred.

Two features of this benchmark depart from a plain repeated game. Utility is transferable within a coalition, so a coalition deviates as a single player: a deviation by S is a leak made jointly by all members of S, who share its value and bear $\kappa(S)$, and a leak supplied by a proper subset $S_1\subsetneq S$ is a deviation of S_1. This attribution is a modeling choice: if tracing exposes only the members whose shares were used, a coalition could leak through a minimal authorized subset and share the gain, so deterrence would have to hold against that subset's exposure. Exposure is charged once, at the leak, and actions other than leaking carry no payoff; Section 5 replaces the one-time charge by probabilistic and time-distributed exposure. The games played after plaintexts become public are outside the model.

4 Main Results

The first theorem gives the exact deterrence threshold in the general reachability model. Appendix A proves the statements of this section in order.

Theorem 4.1 (Exact characterization)**.** *A coalition S is deterred if and only if*

$$\kappa(S)\ge\sup_{t\le T}\Lambda_t(S).$$

Every unauthorized coalition meets this condition, so honest execution is coalition-safe if and only if it holds for every $S\in\mathcal{A}$.

Proof idea. We fix a coalition S and a deviation strategy, and we let τ be the round of its leak. Because the leak is terminal, the buyer reads exactly the rounds in $\mathcal{R}_S(\tau)$, and the exposure is charged once, at τ; so the deviation is worth $\Lambda_\tau(S)-\kappa(S)$ in present value at τ (Lemma A.1). The condition of Theorem 4.1 makes every such value nonpositive, which gives sufficiency. For necessity, a coalition with $\Lambda_t(S)>\kappa(S)$ plays honestly until round t and leaks there. The dynamics enter

only through the choice of τ, and the reachability set records every consequence of the leak, so the theorem needs no model of the market that follows it.

The remaining results answer four design questions: what one-shot reasoning misses (Corollaries 4.2 to 4.4), how large the gap can be (Theorem 4.5 and Proposition 4.6), what bounded reuse buys (Corollary 4.7), and how to split the requirement across parties (Corollaries 4.8 and 4.9). The one-shot rule is exact with a single round and under context binding, while a capability that never expires makes the exposure cover the whole remaining stream.

Corollary 4.2 (One-shot threshold)**.** *In a one-round environment, honest execution is coalition-safe if and only if $\kappa(S) \geq v_1(S)$ for every $S \in \mathcal{A}$.*

Corollary 4.3 (Fully reusable leakage)**.** *Suppose the deployment is fully reusable. Then honest execution is coalition-safe if and only if*

$$\kappa(S) \geq \sup_{t \leq T} \sum_{r=t}^{T} \delta^{r-t} v_r(S) \qquad \text{for every } S \in \mathcal{A}.$$

In particular, if $T = \infty$, $\delta < 1$, and $v_r(S) = v(S)$ for all r, then the exact threshold is $\kappa(S) \geq v(S)/(1-\delta)$.

This requirement grows without bound as $\delta \to 1$.

Corollary 4.4 (Context isolation collapses continuation value)**.** *In a context-isolating deployment, honest execution is coalition-safe if and only if $\kappa(S) \geq \sup_{t \leq T} v_t(S)$ for every $S \in \mathcal{A}$.*

The gap between these two designs can be as large as the horizon.

Theorem 4.5 (Sharp reusable-versus-context-isolating separation)**.** *Let $T \in \mathbb{N}$, and let $S^\star$ be an authorized coalition with $v_t(S^\star) = 1$ for every $t \leq T$ in an instance without discounting ($\delta = 1$). Then:*

(i) under fully reusable leakage, $S^\star$ is deterred if and only if $\kappa(S^\star) \geq T$;

(ii) under context isolation, $S^\star$ is deterred if and only if $\kappa(S^\star) \geq 1$.

Moreover, $\sup_{t \leq T} \sum_{r=t}^{T} \delta^{r-t} v_r(S) \leq T \sup_{t \leq T} v_t(S)$ for every T-round instance and every coalition S, so no instance separates the two thresholds by more than a factor of T.

A committee that budgets one round of value under a long-lived key is short by a factor of T, because one leak collects every round; in every design the gap is driven by how many rounds one leak can reach.

Proposition 4.6 (Reuse multiplicity threshold)**.** *Fix an authorized coalition S, let $\rho(S) = \sup_{t \leq T} |\mathcal{R}_S(t)|$, and suppose $\delta = 1$ and $v_r(S) = 1$ for every round r. Then S is deterred if and only if $\kappa(S) \geq \rho(S)$. Consequently, along a family of deployments in which $\rho(S)$ grows without bound, no exposure that is constant in the horizon deters S.*

A designer who cannot bound $\rho(S)$ cannot certify deterrence with a fixed exposure, while a bound on it gives the closed form below; there $\rho(S)$ denotes any such bound, which may exceed the supremum of Proposition 4.6.

Corollary 4.7 (Bounded reuse multiplicity under discounting)**.** *Fix an authorized coalition S and suppose that $|\mathcal{R}_S(t)| \leq \rho(S)$ for every round t. Assume also that $v_r(S) \leq \bar{v}(S)$ for every round r. Then S is deterred whenever*

$$\kappa(S) \geq \bar{v}(S) \sum_{j=0}^{\rho(S)-1} \delta^j = \begin{cases} \bar{v}(S) \dfrac{1-\delta^{\rho(S)}}{1-\delta}, & \delta < 1, \\ \rho(S)\bar{v}(S), & \delta = 1. \end{cases}$$

Moreover, this bound is tight whenever some round t satisfies $\mathcal{R}_S(t) = \{t, t+1, \ldots, t+\rho(S)-1\}$ and $v_{t+j}(S) = \bar{v}(S)$ for every $j = 0, \ldots, \rho(S)-1$.

Because key rotation every $\rho(S)$ rounds keeps every reachability set within $\rho(S)$ rounds when no leak can include a later key, it caps the multiplier at $\rho(S)$ for every $\delta \leq 1$. For the common equal-stake threshold committee, the general theorem becomes an exact formula.

Setting	$\mathcal{R}_S(t)$	Threshold for coalition S
Context-isolating	$\{t\}$	$\sup_t v_t(S)$
Fully reusable	$\{t, \dots, T\}$	$\sup_t \sum_{r=t}^{T} \delta^{r-t} v_r(S)$
Bounded reuse $\lvert\mathcal{R}_S(t)\rvert \le \rho(S)$, $v_r(S) \le \bar{v}(S)$	arbitrary	at most $\bar{v}(S) \sum_{j=0}^{\rho(S)-1} \delta^j$
Reuse multiplicity $\rho(S)$, unit values, $\delta = 1$	arbitrary	$\rho(S)$
Equal exposure $d\lvert S\rvert$, k-out-of-n	arbitrary	$d^\star = \max_{S \in \mathcal{A}} \sup_t \Lambda_t(S)/\lvert S\rvert$
Additive exposure $\sum_{i \in S} d_i$	arbitrary	choose d_i by the LP of Corollary 4.9

Table 1: Thresholds by reachability pattern: exact where available, sharp bounds otherwise.

Corollary 4.8 (Uniform exposure for k-out-of-n committees)**.** *Suppose $\mathcal{A} = \{S \subseteq [n] : \lvert S\rvert \ge k\}$ for some $1 \le k \le n$ and $\kappa(S) = d\lvert S\rvert$ for a common per-party exposure $d \ge 0$. Then the minimal such d that makes honest execution coalition-safe is*

$$d^\star = \max_{S \in \mathcal{A}} \frac{\sup_{t \le T} \Lambda_t(S)}{\lvert S\rvert}.$$

If moreover $\sup_t \Lambda_t(S)$ depends only on $\lvert S\rvert$, say $\sup_t \Lambda_t(S) = \lambda_{\lvert S\rvert}$, then

$$d^\star = \max_{k \le m \le n} \frac{\lambda_m}{m}.$$

The binding coalition is the size at which reachable value per member peaks, not necessarily the smallest; with unequal amounts the cheapest profile solves a covering program.

Corollary 4.9 (Heterogeneous additive exposure as a covering LP)**.** *Suppose the deployment chooses nonnegative per-party exposures $d_1, \dots, d_n$ and coalition exposure is additive:*

$$\kappa(S) = \sum_{i \in S} d_i.$$

Then honest execution is coalition-safe if and only if $\sum_{i \in S} d_i \ge \sup_{t \le T} \Lambda_t(S)$ for every $S \in \mathcal{A}$. Consequently, the minimum-total-exposure design problem is the linear program

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} d_i \\ \text{s.t.} \quad & \sum_{i \in S} d_i \ge \sup_{t \le T} \Lambda_t(S) && \forall S \in \mathcal{A}, \\ & d_i \ge 0 && \forall i \in [n]. \end{aligned}$$

Theorem 4.1 needs neither additive nor symmetric exposure nor a threshold family $\mathcal{A}$, so it covers heterogeneous slashing and weighted or hierarchical access policies; Table 1 collects the thresholds.

5 Beyond a One-Time Exposure

Two relaxations keep the first-leak argument and change only the cost side of the comparison; both assume risk-neutral coalitions and detection probabilities and fines that the deployment fixes in advance.

Proposition 5.1 (Probabilistic exposure)**.** *Suppose coalition S is exposed after a successful leak with probability $p(S) \in [0, 1]$, and conditional on exposure pays a deterministic fine $f(S) \ge 0$. If the coalition is risk neutral, then it is deterred if and only if $p(S)\, f(S) \ge \sup_{t \le T} \Lambda_t(S)$.*

Tracing mechanisms such as threshold traitor tracing [9] enter only through $p(S)$.

Proposition 5.2 (Time-distributed exposure)**.** *Suppose that if coalition S first leaks at round t, then its expected discounted exposure equals*

$$K_t(S) = \sum_{r \in \mathcal{R}_S(t)} \delta^{r-t} e_{t,r}(S)$$

Input	Source in a deployment	How to bound it
$\mathcal{A}$	decryption policy, such as k of n keypers	read off the protocol
$\mathcal{R}_S(t)$	scope of the leaked object, custody of key shares, key lifecycle	read off the keying design (Table 3)
$v_t(S)$	surplus that early access to round t creates for a buyer	upper bound: replay order flow with and without early access, plus a tail margin
$\kappa(S)$, $K_t(S)$	slashable stake, tracing probability and penalty, forfeited rewards, legal liability	lower bound
δ	discounting per context	$\delta = 1$ over a finite window, or $\delta = e^{-a\Delta}$ for annual rate a and spacing Δ

Table 2: Instantiation map for Theorem 4.1.

for some nonnegative charges $e_{t,r}(S)$. Then S is deterred if and only if $K_t(S) \geq \Lambda_t(S)$ for every $t \leq T$, and honest execution is coalition-safe if and only if this holds for every $S \in \mathcal{A}$.

This covers a fixed charge per compromised round, forfeited future rewards, and evidence whose detection probability grows with continued reuse; Corollary B.1 works out a constant per-round tracing hazard. Random reachability sets and degraded decoders fit the same comparison (Propositions B.2 and B.4).

Long reuse and patient coalitions. If a leaked capability stays valid forever and every round is worth at least $v > 0$ to S, Corollary 4.3 requires exposure at least $v/(1-\delta)$, which diverges as $\delta \to 1$. Two designs keep the requirement bounded. Bounding the lifetime of the leaked object caps the multiplier: when every leak reaches at most $\rho(S)$ rounds, Corollary 4.7 asks for at most $\rho(S)\bar{v}(S)$ for every $\delta \leq 1$. Liability that grows with use works from the cost side: a charge $c \geq \sup_r v_r(S)$ per compromised round gives $K_t(S) = c\sum_{r\in\mathcal{R}_S(t)} \delta^{r-t} \geq \Lambda_t(S)$ at every $\delta < 1$. A fixed one-time loss, reputational or otherwise, does not remove the divergence.

6 Instantiation and Applications

6.1 From a Deployment to the Model

Table 2 lists where each input of Theorem 4.1 comes from; certification needs bounds, not point estimates. For a one-time exposure the comparison in Theorem 4.1 is monotone in each input, so an upper bound on reachable value and a lower bound on exposure certify deterrence; since $\Lambda_t(S)$ is nondecreasing in δ and in $\mathcal{R}_S(t)$, a designer unsure of either can certify at an upper bound. A time-distributed exposure $K_t(S)$ grows with δ and $\mathcal{R}_S(t)$ as well, so its lower bound must hold over the whole range of δ and on the smallest admissible reachability set; a charge of at least $\sup_r v_r(S)$ per compromised round passes for every δ and every reachability set. Competing buyers pay no more than the surplus that early access creates, so $v_t(S)$ also bounds what a coalition can charge. When no feasible exposure clears the bound, the remaining levers are the reach of the leaked object, liability that grows with use (Section 5), the value of early access, and the access structure; under context binding the reach is one round, so per-use liability is one-time and only the last two remain.

6.2 Threshold-Encrypted Mempools

In an encrypted mempool the parties are the keypers or validators that hold key shares, a context is an epoch or a batch, and $v_t(S)$ is what a front-runner would pay for round t's plaintext order flow before the honest release point.

Corollary 6.1 (Encrypted-mempool instantiation)**.** *Consider an encrypted-mempool deployment in which $v_t(S)$ is the maximum transferable-utility gain available to coalition S from selling premature access to round t's plaintext order flow. Then S is deterred in a fully reusable scheme if and only if $\kappa(S) \geq \sup_{t\leq T} \sum_{r=t}^{T} \delta^{r-t} v_r(S)$, and in a context-isolating scheme if and only if $\kappa(S) \geq \sup_{t\leq T} v_t(S)$.*

Table 3 maps keying designs to reachability sets, including two deployed systems. Ferveo runs its distributed key generation again for each validator epoch, and it allows one key to serve more than

Deployment pattern	Leaked object	Reachability $\mathcal{R}_S(t)$
Long-lived threshold key	decoder for the persistent key	$\{t, t+1, \ldots, T\}$
Context-dependent threshold decryption [8], shares in revocable modules	decryption shares bound to χ_t	$\{t\}$
Context-dependent threshold decryption, other custody of shares	a threshold of key shares, or access to modules holding them	epochs up to the next key generation
Rotating committees, fresh key per epoch	decoder for epoch t's key only	$\{t\}$
Windowed key reuse or lagged refresh	decoder valid for w epochs	$\subseteq \{t, \ldots, t+w-1\}$
Ferveo [5], new key each epoch	decryption capability for epoch t	$\{t\}$
Ferveo, one key for L epochs	capability for the shared key	epochs up to the next key generation, at most L
Shutter [16], epoch key leaked	secret key of epoch t	$\{t\}$
Shutter, eon key shares leaked	a threshold of eon key shares, which yield every epoch key	epochs up to the next eon key generation

Table 3: From keying designs to reachability sets. When a coalition can leak several objects, the most valuable one governs (Proposition B.4). Accountability mechanisms that only penalize, such as traitor tracing [9] or accountable logging such as LØ [26], change $\kappa(S)$ or $K_t(S)$ and leave these sets unchanged; revocation or rekeying on detection shrinks them.

Per-attack profit v	$\rho = 1$ (context isolation)	$\rho = 4$	$\rho = 32$ (full reuse)
Median	\$24.07	\$96.28	\$770.24
75th percentile	\$62.92	\$251.68	\$2,013.44
Maximum	\$20,084.01	\$80,336.04	\$642,688.32

Table 4: Required exposure ρv when the per-context value v is set to the Ethereum insertion-attack profits of Torres et al. [33], with $\delta = 1$ over a key lifetime of 32 block contexts.

one epoch at a cost in safety [5]. Shutter's keypers generate a long-lived eon key and derive a key for each epoch from it [16]. A coalition that can produce an epoch key also holds a threshold of eon-key shares, so by Proposition B.4 its exposure must cover the reach of that larger leak, up to the next eon key generation. Committee rotation affects reachability only through key continuity: heavy overlap in membership without key reuse leaves $\mathcal{R}_S(t) = \{t\}$, while a key that stays valid across epochs produces multi-round reuse even with little overlap. Rotation also changes which coalitions hold a threshold: with fresh keys the model sets $v_t(S) = 0$ in epochs where S holds none, and keys that outlive a committee need time-indexed access structures (Section 8).

Example 6.2 (Calibration worksheet)**.** Suppose a conservative cap on one epoch's leakage value is \$50,000 and $\delta = 0.95$. If a leaked decoder stays valid forever, Corollary 4.3 requires $\kappa(S) \geq 50{,}000/(1 - 0.95) = 1{,}000{,}000$. Context isolation reduces this to 50,000 by Corollary 4.4, and a reuse cap of $\rho(S) = 4$ epochs gives $50{,}000(1 + 0.95 + 0.95^2 + 0.95^3) \approx 185{,}494$ by Corollary 4.7. Tracing that succeeds with probability at least 0.4, against a fine of \$500,000 after a trace, gives expected exposure 200,000 by Proposition 5.1. That budget deters context-isolated leakage with room to spare, clears the four-epoch requirement narrowly, and falls five times short of full reuse.

The dollar inputs of Example 6.2 are illustrative, and Table 4 anchors the scale in data. Torres et al. [33] identify 196,691 insertion (sandwich) attacks on Ethereum, with per-attack profit of \$24.07 at the median, \$62.92 at the 75th percentile, and \$20,084.01 at the maximum (their Table 3). We take a key reused over 32 block contexts, set $\delta = 1$ over this short window, and use a constant stage value v; a leak that reaches ρ contexts then requires $\kappa(S) \geq \rho v$. Observed attacker profit is a stress benchmark for $v_t(S)$ and not a certified upper bound, so a deployment certificate should aggregate every profitable response to early access and add tail and estimation margins.

6.3 Other Applications

Federated learning under a collective key. In POSEIDON [29], parties train a neural network under the collective public key of a multiparty homomorphic encryption scheme [25], and decryption requires their cooperation. A training round is a context: a coalition able to decrypt can sell a buyer the round's model or aggregated updates, and $v_t(S)$ is the value of that access. If the collective key persists across rounds, one leak reaches every later round, and refreshing the key per round or per

window of rounds bounds the reachability set as in Corollary 4.7. MedCo [27] applies the same collective-key design to queries on distributed clinical and genomic data, with a query as the context.

Sealed-bid auctions and voting. Sealed-bid auction services keep bids hidden among several servers until the auction closes [17]. When bids are encrypted to a threshold key held by the servers, a coalition that leaks the key's decryption capability opens the bids of every later auction under that key; bids shared afresh for each auction give $\mathcal{R}_S(t) = \{t\}$. In the election scheme of Cramer et al. [13], authorities share a threshold key and decrypt only the tally; a quorum can instead sell individual ballots to a vote buyer, and a key reused across elections lets one leak reach later elections. Ballots are never released, so the leak also opens past elections, which reachability sets do not count; for data that are never released, the threshold of Theorem 4.1 is a lower bound on the exposure needed.

Threshold signatures. The first-leak argument also covers threshold signatures, where the object sold is the ability to act. Then $v_t(S)$ is the total gain from unauthorized signing in context t, and $\mathcal{R}_S(t)$ records the contexts in which a leaked signing capability stays valid: every later epoch for a long-lived bridge key, even when each signature is domain separated, and a single epoch only for independent per-epoch keys or for keys kept in modules that sign only for the current epoch and are replaced as soon as a leak becomes public. Theorem 4.1 applies when forged actions have fixed reachable values; a forged action that changes balances, nonces, or the set of feasible later actions makes values state dependent, which the model does not cover.

7 Limitations and Scope

The inputs $v_t(S)$, $\mathcal{R}_S(t)$, and $\kappa(S)$ are exogenous, and Section 6.1 shows how to bound them.

Theorem 4.1 is exact under transferable utility. When the first-leak benefit and the exposure are deterministic, any increasing coalition utility ranks a deviation by the sign of $\Lambda_t(S) - \kappa(S)$, so risk aversion leaves the threshold unchanged. Risk attitudes matter once exposure or value is random: the expected-exposure reductions of Propositions 5.1 and 5.2 are exact under risk neutrality and with detection that does not respond to later strategic behavior. Without side transfers a coalition has no scalar objective, and the threshold is a transferable-utility benchmark for it.

The benchmark is a leak by one coalition while every other party executes honestly. When authorized coalitions overlap, one coalition's leak can preempt another's; Remark B.8 gives the comparison under exogenous survival probabilities, and the full race is open. The model also takes a leak to be terminal (Section 3.3); when tracing fails to identify the leakers this rests on that assumption alone, since unidentified leakers would receive any replacement keys or modules. A coalition that could keep leaking after a single charge would collect every later round under any keying design, so the fully reusable threshold would apply even under context binding.

Deterring early leakage does not make execution fair after decryption. Proposition B.7 gives a context-isolating deployment in which no coalition leaks and a block builder still sandwiches a user after honest decryption, so ordering after release needs its own mechanism or consensus constraints, and a theory that combines the two layers is open.

8 Open Problems

1. **Endogenous leakage markets.** Derive the stage value $v_t(S)$ from strategic buyers, buyer competition, limited commitment, and competing coalitions.
2. **Dynamic and rotating committees.** Extend the reachability characterization to time-indexed access structures $\mathcal{A}_t$, time-indexed coalitions S_t, and partially overlapping committees.
3. **Partial context isolation.** Corollary 4.7 prices reuse multiplicity alone; exact thresholds for leaked objects that work on a structured subset of future contexts remain open.
4. **Beyond transferable utility.** Which thresholds survive imperfect side transfers, endogenous tracing investment, or risk aversion under random exposure?

Acknowledgments and Disclosure of Funding

This work received no third-party funding. The authors declare no competing interests. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Tropic.

References

[1] Ittai Abraham, Danny Dolev, Rica Gonen, and Joseph Y. Halpern. Distributed Computing Meets Game Theory: Robust Mechanisms for Rational Secret Sharing and Multiparty Computation. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing*, pages 53–62, 2006.

[2] Zeinab Alipanahloo, Abdelhakim Senhaji Hafid, and Kaiwen Zhang. Maximum Extractable Value (MEV) Mitigation Approaches in Ethereum and Layer-2 Chains: A Comprehensive Survey. *IEEE Access*, 12:185212–185231, 2024.

[3] Robert J. Aumann. Acceptable Points in General Cooperative n-Person Games. In A. W. Tucker and R. D. Luce, editors, *Contributions to the Theory of Games IV*, volume 40 of *Annals of Mathematics Studies*, pages 287–324. Princeton University Press, 1959.

[4] Kushal Babel, Nerla Jean-Louis, Yan Ji, Ujval Misra, Mahimna Kelkar, Kosala Yapa Mudiyanselage, Andrew Miller, and Ari Juels. PROF: Protected Order Flow in a Profit-Seeking World. In *2026 IEEE 11th European Symposium on Security and Privacy (EuroS&P)*, pages 398–418, 2026.

[5] Joseph Bebel and Dev Ojha. Ferveo: Threshold Decryption for Mempool Privacy in BFT Networks. *Cryptology ePrint Archive*, Paper 2022/898, 2022.

[6] Gary S. Becker. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, 76(2):169–217, 1968.

[7] B. Douglas Bernheim, Bezalel Peleg, and Michael D. Whinston. Coalition-Proof Nash Equilibria I. Concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.

[8] Dan Boneh, Benedikt Bünz, Kartik Nayak, Lior Rotem, and Victor Shoup. Context-Dependent Threshold Decryption and Its Applications. In *Advances in Cryptology – ASIACRYPT 2025*, Lecture Notes in Computer Science, pages 506–538. Springer, 2025.

[9] Dan Boneh, Aditi Partap, and Lior Rotem. Accountability for Misbehavior in Threshold Decryption via Threshold Traitor Tracing. In *Advances in Cryptology – CRYPTO 2024*, volume 14926 of *Lecture Notes in Computer Science*, pages 317–351. Springer, 2024.

[10] Eric Budish, Andrew Lewis-Pye, and Tim Roughgarden. The Economic Limits of Permissionless Consensus. In *Proceedings of the 25th ACM Conference on Economics and Computation*, 2024.

[11] Vitalik Buterin and Virgil Griffith. Casper the Friendly Finality Gadget. *arXiv preprint arXiv:1710.09437*, 2017.

[12] Yiding Chen, Jerry Zhu, and Kirthevasan Kandasamy. Mechanism Design for Collaborative Normal Mean Estimation. In *Advances in Neural Information Processing Systems 36*, 2023.

[13] Ronald Cramer, Rosario Gennaro, and Berry Schoenmakers. A Secure and Optimally Efficient Multi-Authority Election Scheme. In *Advances in Cryptology – EUROCRYPT '97*, volume 1233 of *Lecture Notes in Computer Science*, pages 103–118. Springer, 1997.

[14] Philip Daian, Steven Goldfeder, Tyler Kell, Yunqi Li, Xueyuan Zhao, Iddo Bentov, Lorenz Breidenbach, and Ari Juels. Flash Boys 2.0: Frontrunning in Decentralized Exchanges, Miner Extractable Value, and Consensus Instability. In *2020 IEEE Symposium on Security and Privacy*, pages 910–927, 2020.

[15] Florian E. Dorner, Nikola Konstantinov, Georgi Pashaliev, and Martin Vechev. Incentivizing Honesty among Competitors in Collaborative Learning and Optimization. In *Advances in Neural Information Processing Systems 36*, 2023.

[16] Stefan Dziembowski, Sebastian Faust, and Jannik Luhn. Shutter Network: Private Transactions from Threshold Cryptography. *Cryptology ePrint Archive*, Paper 2024/1981, 2024.

[17] Matthew K. Franklin and Michael K. Reiter. The Design and Implementation of a Secure Auction Service. *IEEE Transactions on Software Engineering*, 22(5):302–312, 1996.

[18] Joseph Halpern and Vanessa Teague. Rational Secret Sharing and Multiparty Computation: Extended Abstract. In *Proceedings of the Thirty-Sixth Annual ACM Symposium on Theory of Computing*, 2004.

[19] Alireza Kavousi, Duc V. Le, Philipp Jovanovic, and George Danezis. BlindPerm: Efficient MEV Mitigation with an Encrypted Mempool and Permutation. In *29th International Conference on Principles of Distributed Systems (OPODIS 2025)*, volume 361 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 36:1–36:21, 2026.

[20] Kshitij Kulkarni, Theo Diamandis, and Tarun Chitra. Towards a Theory of Maximal Extractable Value I: Constant Function Market Makers. *arXiv preprint arXiv:2207.11835*, 2022.

[21] Rujia Li, Xuanwei Hu, Qin Wang, Sisi Duan, and Qi Wang. Transaction Fairness in Blockchains, Revisited. *IEEE Transactions on Dependable and Secure Computing*, 23(1):752–765, 2026.

[22] George J. Mailath and Larry Samuelson. *Repeated Games and Reputations: Long-Run Relationships*. Oxford University Press, 2006.

[23] Huned Materwala, Shraddha M. Naik, Aya Taha, Tala Abdulrahman Abed, and Davor Svetinovic. Maximal Extractable Value in Decentralized Finance: Taxonomy, Detection, and Mitigation. *IEEE Transactions on Services Computing*, 18(6):4386–4407, 2025.

[24] Peter Bro Miltersen, Jesper Buus Nielsen, and Nikos Triandopoulos. Privacy-Enhancing Auctions Using Rational Cryptography. In *Advances in Cryptology – CRYPTO 2009*, volume 5677 of *Lecture Notes in Computer Science*, pages 541–558. Springer, 2009.

[25] Christian Mouchet, Juan Troncoso-Pastoriza, Jean-Philippe Bossuat, and Jean-Pierre Hubaux. Multiparty Homomorphic Encryption from Ring-Learning-with-Errors. *Proceedings on Privacy Enhancing Technologies*, 2021(4):291–311, 2021.

[26] Bulat Nasrulin, Georgy Ishmaev, Jérémie Decouchant, and Johan Pouwelse. LØ: An Accountable Mempool for MEV Resistance. In *Proceedings of the 24th ACM/IFIP International Middleware Conference*, pages 98–110, 2023.

[27] Jean Louis Raisaro, Juan Ramón Troncoso-Pastoriza, Mickaël Misbach, João Sá Sousa, Sylvain Pradervand, Edoardo Missiaglia, Olivier Michielin, Bryan Ford, and Jean-Pierre Hubaux. MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(4):1328–1341, 2019.

[28] Antoine Rondelet and Quintus Kilbourn. Mempool Privacy: An Economic Perspective. *arXiv preprint arXiv:2307.10878*, 2023.

[29] Sinem Sav, Apostolos Pyrgelis, Juan Ramón Troncoso-Pastoriza, David Froelicher, Jean-Philippe Bossuat, João Sá Sousa, and Jean-Pierre Hubaux. POSEIDON: Privacy-Preserving Federated Neural Network Learning. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2021.

[30] Lloyd S. Shapley. Simple Games: An Outline of the Descriptive Theory. *Behavioral Science*, 7(1):59–66, 1962.

[31] Oliver Stengele, Markus Raiber, Jörn Müller-Quade, and Hannes Hartenstein. ETHTID: Deployable Threshold Information Disclosure on Ethereum. In *Proceedings of the 3rd International Conference on Blockchain Computing and Applications*, pages 127–134, 2021.

[32] Alan D. Taylor and William S. Zwicker. *Simple Games: Desirability Relations, Trading, Pseudoweightings*. Princeton University Press, 1999.

[33] Christof Ferreira Torres, Ramiro Camino, and Radu State. Frontrunner Jones and the Raiders of the Dark Forest: An Empirical Study of Frontrunning on the Ethereum Blockchain. In *30th USENIX Security Symposium*, pages 1343–1359, 2021.

[34] Sarisht Wadhwa, Julian Ma, Thomas Thiery, Barnabe Monnot, Luca Zanolini, Fan Zhang, and Kartik Nayak. AUCIL: An Inclusion List Design for Rational Parties. *Cryptology ePrint Archive*, Paper 2025/194, 2025.

A Proofs for the Main Results

This appendix proves the results of Section 4 in the order in which they are stated.

A.1 Normal Form for Coalition Deviations

The first lemma reduces every coalition deviation to its first successful leak, which turns a repeated deviation into a supremum over first-leak times.

Lemma A.1 (Earliest-leak normal form)**.** *Fix a coalition $S \subseteq [n]$, a round t_0, and a public history with no successful leak before t_0. Consider a pure deviation strategy by S from t_0 on, and let $\tau \geq t_0$ be the round of its successful leak, with $\tau = \infty$ if it never leaks. In present value at round t_0 and relative to remaining honest forever, the deviation yields*

$$\begin{cases} \delta^{\tau - t_0}\big(\Lambda_\tau(S) - \kappa(S)\big), & \text{if } \tau < \infty, \\ 0, & \text{if } \tau = \infty. \end{cases}$$

Proof. We treat the two cases in turn. If $\tau = \infty$, the coalition never leaks, so under the payoff convention of Section 3 it gains and pays nothing.

Suppose $\tau < \infty$. No round before τ is compromised, because τ is the only leak. The leak is terminal, so the buyer reads exactly the rounds in $\mathcal{R}_S(\tau)$, each worth its stage value, and the exposure $\kappa(S)$ is charged once, at τ. Discounting every amount to t_0 gives

$$\sum_{r \in \mathcal{R}_S(\tau)} \delta^{r - t_0} v_r(S) - \delta^{\tau - t_0}\kappa(S) = \delta^{\tau - t_0}\big(\Lambda_\tau(S) - \kappa(S)\big).$$

□

A.2 Exact Characterization

Proof of Theorem 4.1. We fix a coalition S and prove sufficiency first. Suppose that $\kappa(S) \geq \sup_{t \leq T} \Lambda_t(S)$, and take a round t_0, a public history with no leak before t_0, and a pure deviation strategy. By Lemma A.1 the deviation yields 0 or $\delta^{\tau - t_0}(\Lambda_\tau(S) - \kappa(S))$, and the assumed inequality makes the latter nonpositive. A randomized deviation averages these values, so S is deterred.

For necessity, suppose that $\kappa(S) < \sup_{t \leq T} \Lambda_t(S)$. By the definition of the supremum we can choose a round $t^\star$ with $\Lambda_{t^\star}(S) > \kappa(S)$, whether or not the supremum is attained. We let S stay honest until round $t^\star$ and leak there. Evaluated at round 1, this deviation yields, by Lemma A.1,

$$\delta^{t^\star - 1}\big(\Lambda_{t^\star}(S) - \kappa(S)\big) > 0,$$

so S is not deterred. For an unauthorized coalition every reachability set is empty, so $\sup_t \Lambda_t(S) = 0 \leq \kappa(S)$. □

Proof of Corollary 4.2. With $T = 1$ we have $\mathcal{R}_S(1) = \{1\}$ for every authorized S, because $\mathcal{R}_S(1) \subseteq \{1\}$ contains round 1. So $\Lambda_1(S) = v_1(S)$, and Theorem 4.1 gives the claim. □

Proof of Corollary 4.3. We substitute $\mathcal{R}_S(t) = \{t, \dots, T\}$ into Definition 3.4, which gives $\Lambda_t(S) = \sum_{r=t}^{T} \delta^{r-t} v_r(S)$ for every authorized S, and Theorem 4.1 gives the first statement. If in addition $T = \infty$, $\delta < 1$, and $v_r(S) = v(S)$ for all r, the geometric series gives, for every t,

$$\Lambda_t(S) = \sum_{j=0}^{\infty} \delta^j v(S) = \frac{v(S)}{1 - \delta}.$$

□

Proof of Corollary 4.4. Under context isolation we have $\Lambda_t(S) = v_t(S)$ for every authorized S and round t. Theorem 4.1 gives the claim. □

A.3 Separation and Converse Lower Bound

Proof of Theorem 4.5. Under fully reusable leakage with $\delta = 1$ and unit values,

$$\Lambda_t(S^\star) = \sum_{r=t}^{T} 1 = T - t + 1,$$

which is largest at $t = 1$, so $\sup_{t \le T} \Lambda_t(S^\star) = T$ and Theorem 4.1 gives (i). Under context isolation $\Lambda_t(S^\star) = v_t(S^\star) = 1$ for every t, and Theorem 4.1 gives (ii). For the last claim we fix a T-round instance and a coalition S. Since $\delta^{r-t} \le 1$ and $v_r(S) \le \sup_{t' \le T} v_{t'}(S)$, every round $t \le T$ satisfies

$$\sum_{r=t}^{T} \delta^{r-t} v_r(S) \le (T - t + 1) \sup_{t' \le T} v_{t'}(S) \le T \sup_{t' \le T} v_{t'}(S).$$

□

Proof of Proposition 4.6. Since $\delta = 1$ and every stage value of S is 1, each reachable round contributes exactly 1, and we get

$$\Lambda_t(S) = \sum_{r \in \mathcal{R}_S(t)} 1 = |\mathcal{R}_S(t)| \qquad \text{for every round } t.$$

So $\sup_{t \le T} \Lambda_t(S) = \rho(S)$, and Theorem 4.1 makes $\rho(S)$ the threshold for S. Along a family in which $\rho(S)$ grows without bound, this threshold exceeds every constant. □

Proof of Corollary 4.7. We fix a round t. Since $v_r(S) \le \bar{v}(S)$ for every round r and the discount weights δ^{r-t} weakly decrease in the delay $r - t$, a discounted sum over $|\mathcal{R}_S(t)|$ rounds from t on is largest when those rounds come first. Because $|\mathcal{R}_S(t)| \le \rho(S)$, this gives

$$\Lambda_t(S) \le \bar{v}(S) \sum_{j=0}^{|\mathcal{R}_S(t)|-1} \delta^j \le \bar{v}(S) \sum_{j=0}^{\rho(S)-1} \delta^j,$$

and Theorem 4.1 shows that S is deterred. For tightness, suppose that $\mathcal{R}_S(t) = \{t, t+1, \dots, t+\rho(S)-1\}$ for some round t and that each of these rounds has stage value $\bar{v}(S)$. Both inequalities are then equalities at that t, so Theorem 4.1 rules out any smaller exposure. □

A.4 Access-Structure Corollaries

Proof of Corollary 4.8. Since $1 \le k \le n$, the family $\mathcal{A}$ is finite and nonempty and every $S \in \mathcal{A}$ has $|S| \ge 1$, so we may divide by $|S|$. By Theorem 4.1, honest execution is then coalition-safe exactly when

$$d \ge \frac{\sup_{t \le T} \Lambda_t(S)}{|S|} \qquad \text{for every } S \in \mathcal{A},$$

and the smallest such d is the maximum of the right-hand side over $S \in \mathcal{A}$, which is $d^\star$; it is finite by the standing assumption of Definition 3.4. When the numerator depends only on $|S|$, coalitions of equal size impose the same constraint, and the maximum runs over sizes:

$$d^\star = \max_{k \le m \le n} \frac{\lambda_m}{m}.$$

□

Proof of Corollary 4.9. We apply Theorem 4.1 to the additive exposure $\kappa(S) = \sum_{i \in S} d_i$: the deterring profiles are exactly the nonnegative solutions of the covering constraints

$$\sum_{i \in S} d_i \ge \sup_{t \le T} \Lambda_t(S) \qquad \text{for every } S \in \mathcal{A},$$

whose right-hand sides are finite by the standing assumption of Definition 3.4. Minimizing $\sum_i d_i$ over them is a linear program. □

B Extensions, Examples, and Boundary Results

B.1 Proofs of the Exposure Extensions

Proof of Proposition 5.1. We compare with the baseline model in which S bears the deterministic exposure $\kappa'(S) = p(S)f(S)$. The detection probability and the fine are fixed in advance, so a risk-neutral coalition values a deviation by its expected payoff, and the proof of Lemma A.1 values a pure deviation whose leak comes at round τ at $\delta^{\tau-t_0}(\Lambda_\tau(S) - p(S)f(S))$ in expectation at round t_0, its value in that baseline. Theorem 4.1 for κ' gives the claim. □

Proof of Proposition 5.2. We fix a coalition S, a round t_0, and a public history with no leak before t_0. The proof of Lemma A.1, with the expected charge $K_\tau(S)$ in place of $\kappa(S)$, values a pure deviation whose leak comes at round τ at $\delta^{\tau-t_0}(\Lambda_\tau(S) - K_\tau(S))$ and one that never leaks at 0. If $K_t(S) \geq \Lambda_t(S)$ for every t, these values and their averages are nonpositive, so S is deterred. If $K_t(S) < \Lambda_t(S)$ for some t, staying honest until t and leaking there gains $\delta^{t-1}(\Lambda_t(S) - K_t(S)) > 0$ from round 1. Unauthorized coalitions have $K_t(S) = \Lambda_t(S) = 0$. □

B.2 Geometric Evidence Accumulation

Under a constant per-round tracing hazard, the time-distributed exposure of Proposition 5.2 has a closed form.

Corollary B.1 (Geometric evidence accumulation)**.** *Fix an authorized coalition S and let $T = \infty$. Suppose a leak at round t compromises exactly the next $\rho(S)$ rounds, so $\mathcal{R}_S(t) = \{t, t+1, \ldots, t+\rho(S)-1\}$, where $\rho(S)$ may be finite or ∞. Suppose further that each compromised round independently produces fresh tracing evidence with hazard $q(S) \in [0,1]$, and that the coalition pays the fine $f(S)$ upon the first traced round. Then*

$$K_t(S) = \sum_{j=0}^{\rho(S)-1} f(S)\, q(S) \big(\delta(1-q(S))\big)^j.$$

If $\rho(S) < \infty$ and $\delta(1-q(S)) \neq 1$, this equals

$$K_t(S) = f(S)\, q(S)\, \frac{1 - \big(\delta(1-q(S))\big)^{\rho(S)}}{1 - \delta(1-q(S))},$$

and if $\rho(S) = \infty$ and $\delta(1-q(S)) < 1$, then

$$K_t(S) = \frac{f(S)\, q(S)}{1 - \delta(1-q(S))}.$$

Hence S is deterred if and only if this quantity dominates $\Lambda_t(S)$ for every round t.

Proof. We apply Proposition 5.2 with $e_{t,t+j}(S) = f(S)\, q(S)(1-q(S))^j$ for each reachable offset j: the fine $f(S)$ times the probability that the first trace comes j rounds after the leak. The first formula is the resulting sum. Its finite and infinite geometric sums give the closed forms. □

B.3 Stochastic Reuse and Partial Leakage

The same first-leak reduction continues to work if reuse or leakage quality is random, provided the coalition evaluates deviations in expected money terms.

Proposition B.2 (Stochastic reachability)**.** *Suppose that if coalition S leaks at round t, then the compromised rounds form a random set $\widetilde{\mathcal{R}}_S(t) \subseteq \{t, \ldots, T\}$ whose law does not depend on the public history before round t. Define the expected reachable leakage value*

$$\overline{\Lambda}_t(S) = \mathbb{E}\left[\sum_{r \in \widetilde{\mathcal{R}}_S(t)} \delta^{r-t} v_r(S)\right],$$

and assume that it is finite. If the coalition is risk neutral and the expected discounted exposure from a leak at round t is $K_t(S)$, then honest execution is coalition-safe if and only if

$$K_t(S) \geq \overline{\Lambda}_t(S) \qquad \text{for every } S \in \mathcal{A} \text{ and every } t \leq T.$$

Proof. Since the law of $\widetilde{\mathcal{R}}_S(t)$ does not depend on the public history, the expected benefit of a leak at round t is $\overline{\Lambda}_t(S)$ after every history with no earlier leak. We therefore repeat the proof of Proposition 5.2 with $\overline{\Lambda}_t(S)$ in place of $\Lambda_t(S)$: a pure deviation whose leak comes at round τ is worth in expectation, at round t_0,

$$\delta^{\tau-t_0}\big(\overline{\Lambda}_\tau(S) - K_\tau(S)\big).$$

□

Remark B.3. Proposition B.2 covers deployments in which reuse ends at a random time drawn after the leak, such as revocation triggered by stochastic evidence or a memoryless refresh schedule. A refresh schedule that the public history reveals, such as a fixed period with a public phase, belongs to the deterministic model, with $\mathcal{R}_S(t)$ read off the schedule.

Proposition B.4 (Leakage modes and degraded decoders)**.** *For each coalition S, let $\mathcal{L}_S$ be a set of possible leak modes, empty when $S \notin \mathcal{A}$. If mode $\ell \in \mathcal{L}_S$ is chosen at round t, it generates stage values $v_r^\ell(S)$ over a reachable set $\mathcal{R}_S^\ell(t) \subseteq \{t, \dots, T\}$. Define*

$$\Lambda_t^\ell(S) = \sum_{r \in \mathcal{R}_S^\ell(t)} \delta^{r-t} v_r^\ell(S),$$

with a finite supremum over ℓ and t for every $S \in \mathcal{A}$. Then an authorized coalition S is deterred if and only if

$$\kappa(S) \geq \sup_{\ell \in \mathcal{L}_S,\, t \leq T} \Lambda_t^\ell(S),$$

and honest execution is coalition-safe if and only if this holds for every $S \in \mathcal{A}$.

Proof. A leak now consists of a round τ and a mode ℓ, and the proof of Lemma A.1 values a pure deviation that leaks in mode ℓ at round τ at $\delta^{\tau-t_0}(\Lambda_\tau^\ell(S) - \kappa(S))$, evaluated at round t_0. An unauthorized coalition has no mode and never leaks. We then repeat the proof of Theorem 4.1 with the supremum taken over both choices. □

Remark B.5. A degraded decoder that is useful only for a subset of transactions is simply one leakage mode ℓ: its lower coverage appears as smaller stage values $v_r^\ell(S)$, and its weaker portability appears as a smaller reachable set $\mathcal{R}_S^\ell(t)$. The proposition needs no monotonicity or subadditivity assumptions on those objects.

B.4 An Explicit k-out-of-n Example

Example B.6 (Stationary reusable versus context-bound committees)**.** Let $\mathcal{A} = \{S \subseteq [n] : |S| \geq k\}$ for some $1 \leq k \leq n$, let $T = \infty$ and $\delta \in (0, 1)$, and assume that every authorized coalition of size m has constant per-round leakage value $v(m)$. In a fully reusable deployment,

$$\sup_t \Lambda_t(S) = \frac{v(|S|)}{1-\delta},$$

so by Corollary 4.8 the exact equal-exposure requirement is

$$d^\star_{\text{reuse}} = \max_{k \leq m \leq n} \frac{v(m)}{m(1-\delta)}.$$

In a context-isolating deployment, the factor $1/(1-\delta)$ disappears:

$$d^\star_{\text{ctx}} = \max_{k \leq m \leq n} \frac{v(m)}{m}.$$

Thus context isolation removes the entire continuation-value multiplier.

B.5 Proof of the Encrypted-Mempool Corollary

Proof of Corollary 6.1. Under full reuse $\Lambda_t(S) = \sum_{r=t}^T \delta^{r-t} v_r(S)$, and under context isolation $\Lambda_t(S) = v_t(S)$. We apply Theorem 4.1 in each case. □

B.6 What the Model Does Not Solve

The paper's guarantee concerns leakage before the honest release point. It does not imply fairness after plaintexts become public.

Proposition B.7 (Early-leakage deterrence does not imply order fairness)**.** *There exists a deployment in which honest execution is coalition-safe under the exact threshold of Theorem 4.1, yet a block builder can still extract positive value after honest decryption by placing its own transactions around a user's.*

Proof. We take a context-isolating deployment whose exposure satisfies Corollary 4.4, so that every coalition is deterred, and we let one round contain a user order that sells 10 units of token X to a constant-product pool with reserves $(100, 100)$ and no fee, with a minimum output of 8 units of token Y. Every trade preserves the product 10^4 of the reserves. After honest decryption the builder first sells 5 units of X and receives $100/21$ units of Y. The user's order still clears, since it now receives $4000/483 > 8$ units of Y, and the builder then sells its $100/21$ units of Y back for $2645/443$ units of X, a profit of

$$\frac{2645}{443} - 5 = \frac{430}{443} > 0.$$

□

Remark B.8 (Race-to-leak interactions). Theorem 4.1 is a baseline first-leak deterrence statement. A lightweight extension gives coalition S exogenous survival probabilities $\sigma_{t,r}(S) \in [0, 1]$, where $\sigma_{t,r}(S)$ is the probability that S's buyer still enjoys exclusive early access in round r if S leaks at round t. Suppose that S is risk neutral, that access without exclusivity is worth nothing, and that the random set of rounds at which exclusivity survives has a law that does not depend on the public history. Taking that set as the random reachable set, Proposition B.2 with $K_t(S) = \kappa(S)$ gives the comparison $\kappa(S) \geq \sup_t \widetilde{\Lambda}_t(S)$, where

$$\widetilde{\Lambda}_t(S) = \sum_{r \in \mathcal{R}_S(t)} \delta^{r-t} \sigma_{t,r}(S) v_r(S).$$

The hard part is endogenizing the survival probabilities when authorized coalitions overlap and strategically preempt one another. That full race-to-leak equilibrium problem remains outside the present paper.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes].

   Justification: The abstract and introduction state the theorem-level contributions and delimit the model as a stylized theoretical analysis. The main claims are matched to formal definitions, theorem statements, corollaries, and proof sections in the body of the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes].

   Justification: The paper discusses its modeling scope, boundary cases, and open problems in the introduction, model discussion, boundary/extension discussion, and concluding open-problems material. In particular, the analysis is presented as a theoretical characterization rather than an empirical or deployed-system evaluation. Section 7 (Limitations and Scope) collects these limitations in one place.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes].

   Justification: The paper states the assumptions through formal definitions and theorem hypotheses, and provides proofs for the main theorems, lemmas, propositions, and corollaries. Results explicitly marked as sketches or boundary observations are identified as such.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [N/A].

   Justification: The paper does not report experiments. The main results are mathematical statements whose verification depends on the formal model, theorem statements, and proofs provided in the paper.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [N/A].

   Justification: The paper does not use datasets, trained models, or experimental code to support its claims. Algorithmic statements, where present, are specified and analyzed mathematically in the text.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [N/A].

   Justification: The paper contains no training, test set, hyperparameter, optimizer, or empirical evaluation setup. Its results are derived from the stated theoretical model.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [N/A].

   Justification: The paper does not include experiments or statistical estimates. Consequently, there are no empirical error bars, confidence intervals, or significance tests to report.

   Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [N/A].

   Justification: The paper does not run computational experiments. No experimental compute resources are required to reproduce the paper's main claims.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes].

   Justification: The work is a theoretical analysis and does not involve human subjects, private data, deployed interventions, or release of potentially harmful models or datasets. The paper preserves anonymity in the submission version.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes].

Justification: The paper discusses positive implications for robust, privacy-preserving, or incentive-compatible decentralized mechanisms, while also identifying boundary cases and residual attack surfaces. The work is theoretical and does not propose a deployed system, so the impact discussion is correspondingly scoped to technical security and market-design implications.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A].

Justification: The paper does not release data, trained models, scraped datasets, or other assets with high misuse risk. No release-specific safeguards are therefore applicable.

Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [N/A].

Justification: The paper does not use existing code, datasets, models, benchmarks, or other external assets as research inputs. Prior scholarly work is credited through citations in the related-work and references sections.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [N/A].

Justification: The paper does not introduce or release new datasets, code packages, models, or benchmarks. Its contribution consists of formal models, theorems, and proofs.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A].

Justification: The paper does not involve crowdsourcing, surveys, experiments with human participants, or human-subject data collection. There are therefore no participant instructions, screenshots, or compensation details to report.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A].

    Justification: The paper does not involve human subjects or crowdsourced participants. IRB or equivalent human-subjects review is therefore not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes].

    Justification: AI tools were used to help check the manuscript's derivations and proofs, and to provide suggestions for improving and revising the derivations.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.