Legitimate Overrides in Decentralized Protocols

Oghenekaro Elem[1] and Nimrod Talmon[2]

[1]Parametrig, karo@parametrig.com
[2]BGU, IOG, talmonn@bgu.ac.il

Abstract

Decentralized protocols claim immutable, rule-based execution, yet many embed emergency mechanisms such as chain-level freezes, protocol pauses, and account quarantines. These overrides are crucial for responding to exploits and systemic failures, but they expose a core tension: when does intervention preserve trust and when is it perceived as illegitimate discretion? With approximately \$10 billion in technical exploit losses potentially addressable by onchain intervention (2016–2026), the design of these mechanisms has high practical stakes, but current approaches remain ad hoc and ideologically charged. We address this gap by developing a *Scope* $\times$ *Authority* taxonomy that maps the design space of emergency architectures along two dimensions: the precision of the intervention and the concentration of trigger authority. We formalize the resulting tradeoffs of a standing centralization cost versus containment speed and collateral disruption as a stochastic cost-minimization problem; and derive three testable predictions. Assessing these predictions against 705 documented exploit incidents, we find that containment time varies systematically by authority type; that losses follow a heavy-tailed distribution ($\alpha \approx 1.33$) concentrating risk in rare catastrophic events; and that community sentiment measurably modulates the effective cost of maintaining intervention capability. The analysis yields concrete design principles that move emergency governance from ideological debate towards quantitative engineering. An extended version of this paper with detailed incident analysis, supplementary figures, and expanded discussion is available here.

1 Introduction

Hacks and exploits are a persistent feature of blockchains and DeFi protocols, repeatedly producing losses that are large relative to protocol treasuries and TVL. E.g., according to Charoenwong and Bernardi (2021; revised 2025) [12], cumulative losses from protocol failures, exploits, and market manipulation approaches \$88 billion. While much of this value derives from systemic market failures (e.g., Terra/Luna), a significant persistent strata ($\approx$ \$10 billion) consists of technical exploits potentially addressable by onchain emergency mechanisms.

A canonical early episode is the 2016 *DAO* exploit [20], whose aftermath culminated in a socially coordinated chain reconfiguration (the Ethereum hard fork) [21], illustrating that "immutability" is ultimately mediated by governance when stakes are high. In response, many systems embed *emergency mechanisms* intended to limit damage under time pressure: protocol- or module-level pauses and shutdown procedures, issuer- or contract-level blacklisting/freeze controls, and (in some networks) chain-level transaction restrictions. Such mechanisms can be effective in containment, but they introduce a legitimacy and centralization tension: intervention power creates an additional attack/abuse surface and changes the system's trust model; moreover, even when never exercised, the mere *existence* of privileged override capability may reduce perceived trustlessness and thus depress utility or valuation.

Remark 1. *Consistent with this, a recent large-scale scan of 166 blockchain networks reports that 16 chains contain active fund-freezing functions and another 19 could introduce similar capabilities with relatively minor changes (35/166 ≈ 21%) [11]; the same report highlights prominent deployments of these capabilities in practice, including the use of* hardcoded *blacklists (addresses embedded directly in the chain's configuration, as used during the $570M BNB Chain bridge exploit [10]) and* config-based *mechanisms (runtime-configurable lists, as deployed in the $162M freeze of stolen assets on Sui after the Cetus hack [47]).*

Immutability–intervention paradox. Emergency mechanisms are introduced to prevent catastrophic safety failures, yet they also create a second-order governance risk: they alter the trust model by introducing privileged discretion (or privileged *optionality*) over state transitions. This yields an immutability–intervention paradox: in crises, communities often *demand* intervention to protect users and integrated systems, but outside crises, the same intervention capability can be viewed as a standing centralization backdoor whose very existence reduces credibility.

Safety–liveness tradeoff (SLT). We use the standard distributed-systems distinction: *safety* prohibits "bad" state transitions (e.g., theft or invalid state), whereas *liveness* guarantees that "good" progress eventually happens (e.g., transactions continue to be processed). Emergency overrides typically improve safety by restricting transitions, but thereby reduce liveness (and often censorship resistance). Here, we are mainly interested in quantifying this tradeoff towards optimizing emergency mechanisms for different settings.

Contributions. These are our main contributions:

- **Design-space taxonomy.** We propose a compact *Scope* × *Authority* taxonomy that organizes emergency governance architectures into a single design space.
- **Incident mapping.** We document prominent override episodes (including chain-level responses and governance-led reconfigurations) and use them to "fill" the taxonomy, highlighting legitimacy tensions that recur across ecosystems.
- **Decision support framework.** We formalize the trade-offs between standing centralization and blast radius into a quantitative cost model and provide an open-source *Intervention Mechanism Calculator*. This tool enables protocol designers to calibrate mechanism selection based on community sentiment and estimated threat probabilities.[1]

[1]The tool is available here.

Roadmap. Section 2 positions our work within relevant literatures. Section 3 presents our Scope × Authority taxonomy. Section 4 formalizes the decision problem as a stochastic optimization. Section 5 assesses the model with comprehensive empirical analysis. Section 6 extracts design principles for practitioners.

2 Related Work

Our work intersects four bodies of literature.

Blockchain security and incident response. The security community has documented exploit patterns and proposed defensive mechanisms [12]. Recent comprehensive reviews by Dwivedi et al. [17] and Siam et al. [41] systematically categorize vulnerabilities across blockchain layers – from network-level attacks to smart contract exploits – yet most work focuses on documentation and *ex-post* analyses rather than prevention, preemption, or post-incident governance. Notable exceptions include analyses of the DAO fork [9] and studies of bug bounty programs as coordinated disclosure mechanisms [5]. Importantly, *legitimate overrides* and the broader design of formal intervention mechanisms remain significantly under-studied in both academic and practitioner literature; our work aims to fill this gap by providing a quantitative framework.

Pre-execution prevention primitives. An emergent class of mechanisms operates *before* transaction execution, shifting the intervention point from reactive pauses to deterministic pre-execution enforcement. The Phylax Credible Layer [36] exemplifies this paradigm: protocols define "assertions" – Solidity-written invariants that specify states which should never occur (e.g., "a healthy account cannot become liquidatable in a single transaction"). The network's sequencer then simulates every transaction against registered assertions and drops violations before execution. This architecture inherits the chain's trust assumptions (no external oracle or monitoring service) and provides zero false positives by construction. Early adopters include Euler Finance ("Holy Grail" invariant for account liquidity) [23], Malda Protocol, Turtle/Lagoon, and Denaria, each deploying 1–6 assertions to protect critical protocol invariants. Linea's integration of the Credible Layer into its sequencer (January 2026) [36] demonstrates institutional appetite for infrastructure-level security guaranties. This primitive occupies a distinct cell in our taxonomy: Network Scope (sequencer-level enforcement) with Signer Set Authority (sequencer operator controls assertion enforcement), offering a compelling alternative to reactive pauses in risk-averse environments. Unlike reactive mechanisms that respond after exploit detection, pre-execution prevention operates deterministically at the transaction level before execution, representing a fundamentally different paradigm that extends beyond our core 5×3 framework of reactive interventions (which operate at network, asset, protocol, module, or account scope).

DAO governance. Recent empirical work examines onchain governance mechanisms, voting behavior, and the tension between efficiency and decentralization [50, 34, 51]. Wang et al. [50] analyzed 581 DAOs and 16,246 proposals to document governance dynamics, while Ma et al. [34] examined 3,348 DAOs across 9 blockchains, revealing widespread governance vulnerabilities including contract backdoors and malicious proposals. Qian [38] demonstrated how flash loan attacks, offchain voting manipulation, and token-based coercion led to over $300M in losses across major DAOs. Werbach et al. [51] surveyed 23 blockchain projects to compare onchain and offchain governance practices, highlighting the gap between pristine abstractions and messy realities. Our focus on *emergency* governance-decisions under time pressure complements this literature.

Political science of emergency powers. The political science literature on states of emergency, constitutional constraints, and executive overreach provides conceptual tools for analyzing blockchain emergency mechanisms. Scholars have long studied how democracies balance speed and accountability during crises, a tension that maps directly to our Authority dimension [24, 27]. We bridge this connection to open a dialogue between blockchain governance and formal constitutional theory of exception.

3 Taxonomy: Scope $\times$ Authority

Our central organizing device is a two-dimensional taxonomy of emergency mechanisms along two orthogonal design dimensions: (i) *Scope* (the precision / blast radius of the intervention), and (ii) *Authority* (who can trigger it, and how). The safety–liveness profile is treated as an induced property of these design choices, rather than an axis in itself.

3.1 Dimension I: Scope (Hierarchy of Precision)

We model scope as a discrete hierarchy of precision levels. Moving downward increases precision (reduces blast radius) and typically reduces collateral disruption, but may require stronger instrumentation and may increase response complexity.

1. **Network scope.** Chain-wide restriction or reconfiguration affecting all applications (e.g., halt/pause, chain-wide censorship rules, reorg/rollback, global fork-based remediation).

2. **Asset scope.** Actions targeting a specific asset across holders/venues (e.g., issuer blacklisting, asset-specific freezes/burns, bridge-wide caps on a given token).

3. **Protocol scope.** Application-wide restriction within a specific protocol (e.g., pausing all markets in a lending protocol; emergency shutdown of a stablecoin system).

4. **Module scope.** Feature-specific restriction within a protocol (e.g., pausing liquidations while allowing deposits/repayments; pausing oracle updates).

5. **Account scope.** Targeted restriction or remediation affecting specific addresses/accounts only (e.g., freezing or quarantining addresses implicated by evidence).

3.2 Dimension II: Authority (Trigger Holder)

We distinguish three authority modes, ordered from concentrated to broadly distributed:

1. **Signer set (key-based).** A fixed keyholder set (e.g., 1-of-n or m-of-n multisig) can trigger the mechanism.
2. **Delegated body.** A designated council/committee (typically multi-party) holds bounded emergency powers, often with mandates, reporting duties, and ex post accountability.
3. **Governance process.** The intervention requires a formal vote or a broadly coordinated social process (e.g., token-holder governance, validator/community coordination for upgrades).

4 A Stochastic Model of Emergency Governance

The Scope $\times$ Authority taxonomy (Section 3) suggests that emergency governance is, fundamentally, a *design problem under time pressure and uncertainty.* A protocol must choose ex ante which override capabilities to embed (if any), how precise they can be, and who can trigger them; then, when an incident occurs, the chosen architecture constrains which responses are feasible and how quickly they can be executed. The empirical cases in Section B demonstrate that similar threat events yield very different interventions, and that perceived legitimacy depends not only on outcomes but also on scope, authority, and procedural safeguards.

Design objective (informal). At a high level, an emergency mechanism trades off three ingredients: (i) *containment* of losses from an unfolding incident (which typically favors fast, powerful triggers), (ii) *collateral disruption* to uninvolved users and applications (which typically favors higher precision), (iii) the *standing centralization cost* of privileged override capability (which is incurred even when no emergency occurs, since it changes the trust model and increases perceived discretion).

4.1 A Minimal Stochastic Model

A protocol designer chooses an emergency governance architecture $m \in \mathcal{M}$. Future adverse events are modeled by a finite set of types $\mathcal{H}$ and a distribution $\Pr[\cdot]$ over $\mathcal{H}$ (optionally including a "no-incident" type; think of it as a static decision for the next timestep).

If an event of type $h \in \mathcal{H}$ occurs and is not yet contained, it generates loss at rate $\mathrm{DamageRate}(h) \geq 0$ per unit time (the "no-incident" type has $\mathrm{DamageRate}(h) = 0$). Architecture m contains the event after $\mathrm{Time}(m) \geq 0$ time units. Exercising m also induces a collateral disruption cost $\mathrm{BlastRate}(m) \geq 0$ (capturing the fixed, one-time shock of its scope). Finally, m imposes a standing centralization cost $\mathrm{CentralizationCost}(m) \geq 0$, incurred regardless of whether an incident occurs (Chekhov's gun).

Scope \ Authority	Signer set	Delegated body	Governance process
Network	*Key-triggered chain-wide restriction* (e.g., halt/pause; global censorship toggle) **Examples:** Harmony Horizon (Jun 2022, $100M) [30]; BNB Chain halt (Oct 2022, $570M) [39]; Berachain halt (Nov 2025) [8]	*Council-coordinated network action* (e.g., bounded emergency council mandate) **Examples:** Poly Network validator coordination (Aug 2021, $611M returned) [19]	*Governance-led chain reconfiguration* (e.g., coordinated upgrade/fork-based remediation) **Examples:** Ethereum DAO fork (Jun 2016) [20]; Gnosis Chain hard fork (Dec 2025, $9.4M) [29]
Asset	*Issuer/admin asset controls* (e.g., blacklist/freeze/burn hooks) **Examples:** Tether USDT freezes (PDVSA $182M) [52]; Circle USDC blocked addresses (Tornado Cash $75K) [13]; WLFI blacklist of Justin Sun ($107M) [14]	*Delegated asset committee* (e.g., bridge/operator council enforces caps/blocks) **Examples:** Gnosis Bridge Board freeze (Nov 2025) [28]; Curve Emergency DAO (emissions/PSR only) [33]	*Governance changes asset rules* (e.g., parameter change, migration, social recovery) **Examples:** Yearn governance-approved pxETH burn [53]
Protocol	*Admin pause / shutdown* (e.g., protocol-wide circuit breaker) **Examples:** Balancer CSPv6 auto-pause (Nov 2025) [7]; Liqwid Kill Switch [32]; Beanstalk shutdown (Apr 2022, $182M); Superfluid agreement halt (Feb 2022)	*Security-council pause* (e.g., bounded emergency mandate) **Examples:** Aave Protocol Guardians [2]; Balancer V3 Emergency subDAO [6]; Liqwid Pause Guardian (4-of-X)	*DAO-administered emergency action* (e.g., vote to pause/upgrade/settle) **Examples:** MakerDAO ESM (deprecated) [42]; Anchor Protocol (defunct) [4]
Module	*Admin disables a feature* (e.g., stop liquidations / withdrawals) **Examples:** Compound price oracle pauses (2021); FEI/Rari DAI borrow pause (Apr 2022, $80M); Elephant Money TRUNK minting pause (Apr 2022)	*Delegated feature-specific pauses* (e.g., guardian pauses liquidations) **Examples:** Aave V3 reserve/pool pauses [3]; Liqwid market pause (Oct 2025); dYdX YFI circuit breaker [18]	*Governance toggles module parameters* (e.g., vote to enable / disable a feature temporarily) **Examples:** MakerDAO (Sky) USDC-PSM pause (Mar 2023) [44]; Aave asset parameter updates [1]; Solend USDH LTV set to 0 (Nov 2022)
Account	*Key-based targeted restriction* (e.g., freeze/quarantine addresses) **Examples:** Tether/Circle address blacklists; Sonic freezeAccount (Nov 2025) [45]; Ronin attacker freeze [43]	*Delegated targeted remediation* (e.g., council-authorized quarantines) **Examples:** StakeWise multisig burn/mint ($20.7M) [46]; Flow Isolated Recovery (Dec 2025, 1,060 addresses) [25]	*Governance-authorized targeted action* (e.g., vote-based remediation against identified addresses) **Examples:** Sui/Cetus 90.9% stake vote (May 2025, $162M) [47]; VeChain blocklist (Dec 2019) [49]

Table 1: **Emergency mechanisms mapped to Scope (precision) $\times$ Authority (trigger holder).** The table defines the design space used to structure the narrative evidence and, later, the formal analysis. Some incidents span cells (e.g., Sui/Cetus: Delegated Body freeze $\rightarrow$ Governance recovery vote).

Thus, given a distribution over bad-event types $h \in \mathcal{H}$ – specified by their probabilities $\Pr[h]$ and damage rates $\mathrm{DamageRate}(h)$ – and given an emergency governance architecture $m \in \mathcal{M}$ with containment time $\mathrm{Time}(m)$, standing centralization cost $\mathrm{CentralizationCost}(m)$, and blast-radius cost $\mathrm{BlastRate}(m)$, the designer's task is to minimize the expected cost. We define the expected cost of architecture m by

$$\begin{aligned}\mathrm{ExpectedCost}(m) \;:=\; & \mathrm{CentralizationCost}(m) \\ & + \sum_{h \in \mathcal{H}} \Pr[h] \cdot \Big(\mathrm{Time}(m) \cdot \mathrm{DamageRate}(h) + \mathrm{BlastRate}(m)\Big),\end{aligned}$$

and study the design problem

$$\min_{m \in \mathcal{M}} \mathrm{ExpectedCost}(m).$$

4.2 Three Theoretical Predictions

Our model yields three testable predictions that we assess in Section 5:

1. **Prediction 1 (Speed-Centralization Tradeoff):** Faster architectures (Signer Set) minimize exploit losses but impose higher standing centralization costs than slower architectures (Governance).

2. **Prediction 2 (Scope-Blast Relationship):** Higher-precision interventions (Account, Module scope) achieve comparable containment outcomes with substantially lower collateral disruption than broader interventions (Protocol, Network scope).

3. **Prediction 3 (Sentiment-Cost Modulation):** Positive community sentiment toward emergency mechanisms reduces their effective standing centralization cost; negative sentiment increases it.

5 Empirical Analysis and Assessment

Building on the theoretical framework in Section 4, we now present comprehensive empirical evidence using data from 705 documented exploit incidents (2016–2026). Motivated by Pearson's Law that *"that which is measured and reported improves exponentially,"* we develop measurement frameworks to illuminate how legitimate emergency interventions can reduce protocol losses and protect users.

5.1 Dataset Overview and Stratification

A critical distinction in our analysis is the separation between "systemic failures" and "intervention-eligible exploits." We stratify the 705 documented cases into four categories:

1. **Systemic Failures** (10 cases, \$61.80B): Massive economic design collapses (e.g., Terra/Luna, FTX) where no emergency pause mechanism could prevent loss.

2. **Other Non-Addressable** (94 cases, \$7.41B): Incidents like rug pulls, phishing, or unpausable logic bugs where intervention was not technically feasible.

3. **Intervention-Eligible** (601 cases, \$9.60B): Technical exploits (reentrancy, logic bugs, etc.) where emergency mechanisms were applicable.

4. **Actually Intervened** (130 cases, \$7.51B): Cases where emergency mechanisms were actually activated.

Our effectiveness analysis focuses strictly on the **601 intervention-eligible cases**. Including systemic failures (as done in raw aggregators) would severely distort the analysis, as a \$40B collapse like Terra is not "addressable" by the emergency governance mechanisms we study.

5.2 Loss Distribution: Power Law Validation

The resulting distribution of losses (Figure 4) follows a power law (Kolmogorov-Smirnov test statistic $D = 0.150$, $p < 0.001$, Power Law exponent $\alpha \approx 1.33$), confirming that risk is driven by fat-tail events. This Pareto-like pattern (approximately 80% of cumulative losses from fewer than 50 incidents) implies that the expected value of intervention capability is driven primarily by its effectiveness against "super-hacks," where rapid containment can prevent tens or hundreds of millions in additional losses.

5.3 The Speed-Centralization Tradeoff

Our intervention incidents data (Figure 7, Figure 1) confirms the model's Prediction 1: containment time Time(m) varies systematically by authority type. For the 52 verified intervention incidents:

- **Signer Set** interventions achieve median containment in approximately 30 minutes.
- **Delegated Body** interventions require approximately 60–90 minutes.
- **Governance** interventions, when they occur, operate on timescales of days to weeks (e.g., the Gnosis hard fork required ≈30 days).

This ordering aligns with intuition: concentrated authority enables faster response, while distributed authority introduces coordination latency.

5.4 Speed–Effectiveness Tradeoff

Figure 2 visualizes the relationship between reaction speed (Time(m)) and containment success for 52 high-fidelity case studies. We define Time(m) using two observable timestamps: (i) *time-to-detect* = minutes between the first credible alert and the first containment trigger, and (ii) *time-to-contain* = minutes between detection and mechanism execution (pause/freeze/halt). The data supports Prediction 1: faster architectures (like *Signer Set*) significantly minimize loss, though they carry higher centralization costs.

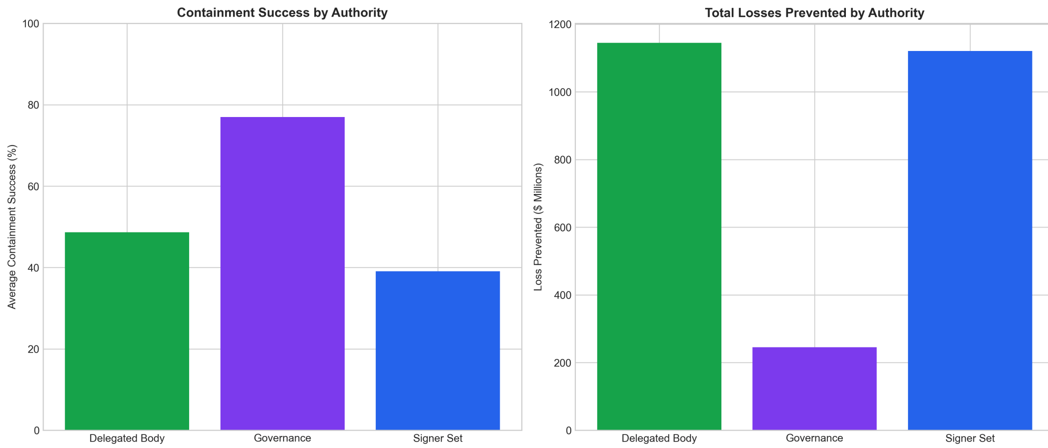

Figure 1: **Intervention Success Rates.** Comparison of containment success across authority types. Signer Set interventions (left) show higher reliability in halting exploits compared to Delegated Bodies, likely due to reduced coordination latency. Right panel shows total losses prevented by authority type.

5.5 Community Sentiment: Calibrating Centralization Cost

We collected governance forum discussions (Discourse, Twitter) around intervention incidents and applied an automated sentiment analysis pipeline. For each incident, we extracted $k = 20$ representative posts (where available) and processed them using VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon-based analyzer [31]. VADER is optimized for social media and forum text as it accounts for both lexical features and structural signals such as punctuation and intensifiers. The analyzer returns a normalized *compound* score $s \in [-1, 1]$, where $+1$ is maximally positive and -1 is maximally negative.

The aggregate average sentiment across 271 posts was $\mathbf{+0.028}$, suggesting that in practice, communities tend to accept emergency interventions, though with significant variance.

This validates Prediction 3 and our standing centralization cost formulation:

$$\text{CentralizationCost}(m) = \text{MarketCap} \times \text{DiscountRate}(m) \times (1 - \bar{s})$$

where $\bar{s} \in [-1, 1]$ is the average sentiment score. Positive sentiment reduces the effective discount rate (community accepts the mechanism), while negative sentiment increases it (community views mechanism as overreach).

Our 52 verified cases reveal a paradox: Signer Sets dominate volume (37 cases, 71.2%) but achieve only 39.1% success, while Governance achieves 73.2% success on its smaller subset (6 cases). This reframes the “Immutability Paradox” as a *Scope-Authority Matching* problem: protocols fail not because governance is “slow,” but because they deploy Direct Democratic mechanisms for Oligarchic tasks (fast containment) while underutilizing them for Constitutional tasks (network recovery).

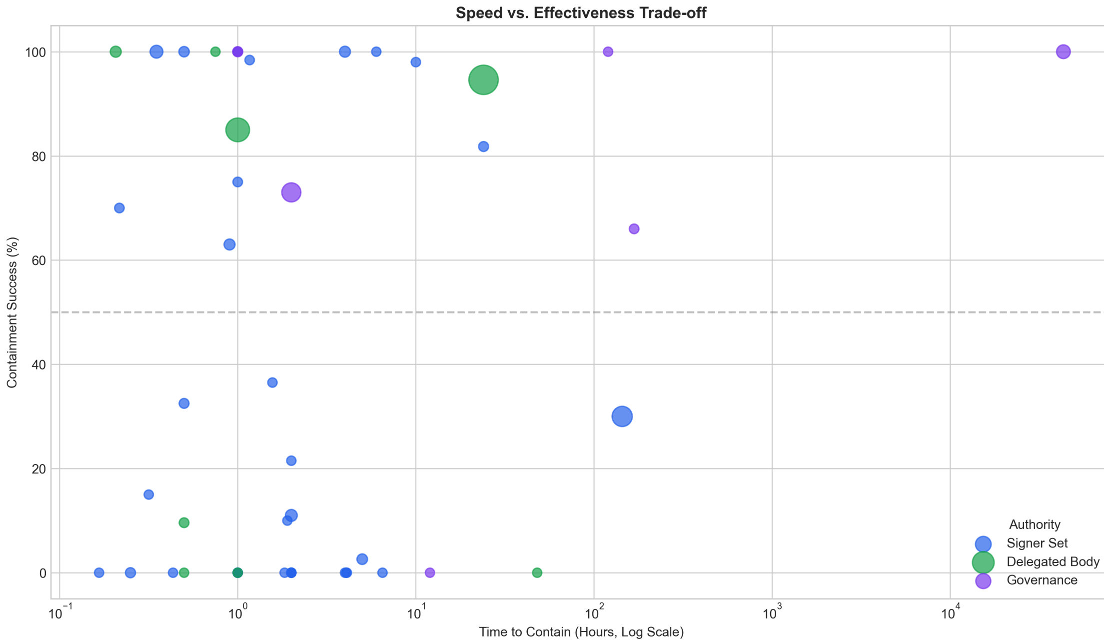

Figure 2: **Speed-Effectiveness Trade-off.** Relationship between time-to-containment (log scale, hours) and loss prevented. Faster interventions (left side) consistently preserve more value, empirically validating the model's containment term. Bubble size represents loss prevented.

6 Design Implications

Our theoretical model (Section 4) and empirical findings (Section 5) jointly suggest several design principles for emergency governance in decentralized protocols.

6.1 The Delegation Sweet Spot

The expected cost function $\text{ExpectedCost}(m)$ reveals a non-monotonic relationship between authority concentration and total social cost. Our empirical Speed-Scope-Success Paradox data validates this theoretical prediction:

- **Pure governance** (m_{gov}) minimizes standing centralization cost but maximizes containment time, making it unsuitable for fast-moving exploits. Empirically, this achieves 73.2% success but only on 11.5% of cases (Network-scope interventions).
- **Signer set** (m_{key}) minimizes containment time but imposes a large standing trust tax, reducing protocol valuation even absent any incident. Empirically, this dominates volume (71.2% of cases) but achieves only 39.1% containment success.
- **Delegated body** (m_{council}) occupies the empirical sweet spot: 48.6% success rate on 15.4% of cases with \$0.88B protected, reflecting bounded authority with faster-than-governance response but without the Signer Set's trust tax.

The emergence of *Emergency subDAOs* (pioneered by Curve, adopted by Balancer V3, and variants like Aave's Protocol Guardians) reflects practitioner convergence toward this sweet spot. These bodies operate under explicit mandates, time-bounded powers, and reporting requirements, features that reduce CentralizationCost(m) while preserving containment speed.

6.2 Scope Matters: Precision Reduces Blast Radius

Our taxonomy distinguishes five levels of scope, from network-wide halts to account-level freezes.

The empirical data shows that *higher-precision interventions* (Asset and Account scope) achieve comparable containment outcomes with substantially lower collateral disruption. For protocol designers, this implies:

- Invest in *instrumentation* that enables targeted response (e.g., per-account freeze hooks, module-level circuit breakers).
- Avoid reliance on "nuclear options" (full chain halts) except as a last resort.
- Design upgrade paths that *increase* precision over time as tooling matures.

6.3 The Culture Multiplier

Blast radius cost is not uniform across ecosystems. Chains with strong "DeFi-maxi" or permissionless cultures suffer disproportionate reputational damage from interventions, even successful ones.

Conversely, chains targeting regulated use cases (RWA, payments, institutional custody) may experience a *regulatory premium*: the presence of robust override capability increases rather than decreases valuation.

Designers should calibrate their mechanism choice to community expectations.

A "culture multiplier" γ can be incorporated into the model:

$$\text{BlastRate}(m) = \gamma \cdot \text{Scope\%} \cdot \frac{\text{Daily Volume}}{1440}$$

where γ is high for permissionless chains and low for compliance-oriented chains.

6.4 A Decision Framework

To make the theoretical model actionable, we map observable protocol and exploit parameters to the formal model variables in Table 2. This allows designers to calibrate the calculator tool using evidence-based heuristics.

We propose that protocol designers use our stochastic model as a decision support tool:

1. **Estimate threat parameters**: Probability distribution $\Pr[h]$ and damage rates DamageRate(h) for plausible exploit scenarios.
2. **Estimate mechanism costs**: Standing centralization cost CentralizationCost(m), containment time Time(m), and blast rate BlastRate(m) for candidate architectures.

Table 2: **Mapping Protocol Parameters to Model Variables.** This table standardizes the mapping between observable protocol and exploit characteristics and formal variables in our expected cost model.

Parameter	**Relevance**	**Model Variable**	**Short Explanation**
Protocol Type	Asset risk profile	$\Pr[h]$, $\mathrm{DamageRate}(h)$	AMM vs Lending vs Bridge risk profiles
Exploit Type	Containment urgency	$\mathrm{DamageRate}(h)$	Flash loan (fast) vs Reentrancy (medium)
Exploit Novelty	Variant vs. Zero-day	$\Pr[h]$, $\mathrm{DamageRate}(h)$	Zero-day = max damage rate
Audit Status	Preventive security health	$\Pr[h]$	Multiple audits = lower threat prob
Community Sentiment	Political legitimacy	$\mathrm{CentralizationCost}(m)$, $\mathrm{BlastRate}(m)$	Trust reduces political cost
TVL Affected	Economic scale at risk	$\mathrm{BlastRate}(m)$	Higher TVL = larger blast radius
Security Claims	Breach accountability	$\mathrm{CentralizationCost}(m)$	"Immutable" = high trust tax

3. **Minimize expected cost**: Choose the architecture m^* that minimizes $\mathrm{ExpectedCost}(m)$ given the protocol's risk profile and community culture.

This framework moves emergency governance from ideology ("decentralization good, admin keys bad") to quantitative cost-benefit analysis.

7 Conclusion

Emergency override mechanisms are a pervasive yet under-theorized feature of decentralized protocols. This paper has offered three complementary contributions toward filling that gap. First, we introduced a compact *Scope* × *Authority* taxonomy that organizes the heterogeneous landscape of emergency mechanisms into a single, two-dimensional design space, and we populated it with prominent real-world episodes spanning chain-level halts, asset freezes, protocol pauses, module toggles, and account-level quarantines. Second, we formalized the underlying design tradeoff as a stochastic cost-minimization problem that balances standing centralization cost, containment speed, and collateral disruption—and derived three testable predictions from the model. Third, we assessed these predictions against 705 documented exploit incidents (2016–2026), finding that containment time varies systematically by authority type, that losses follow a power-law distribution concentrating risk in rare "super-hacks," and that community sentiment measurably modulates the effective cost of intervention capability. Together, these results reframe emergency governance as a quantitative engineering discipline rather than an ideological binary, and provide actionable design principles, delegation sweet spots, precision instrumentation, culture-aware calibration, and conditional sunset clauses, for protocol designers navigating the immutability–intervention paradox.

A Supplementary Figures

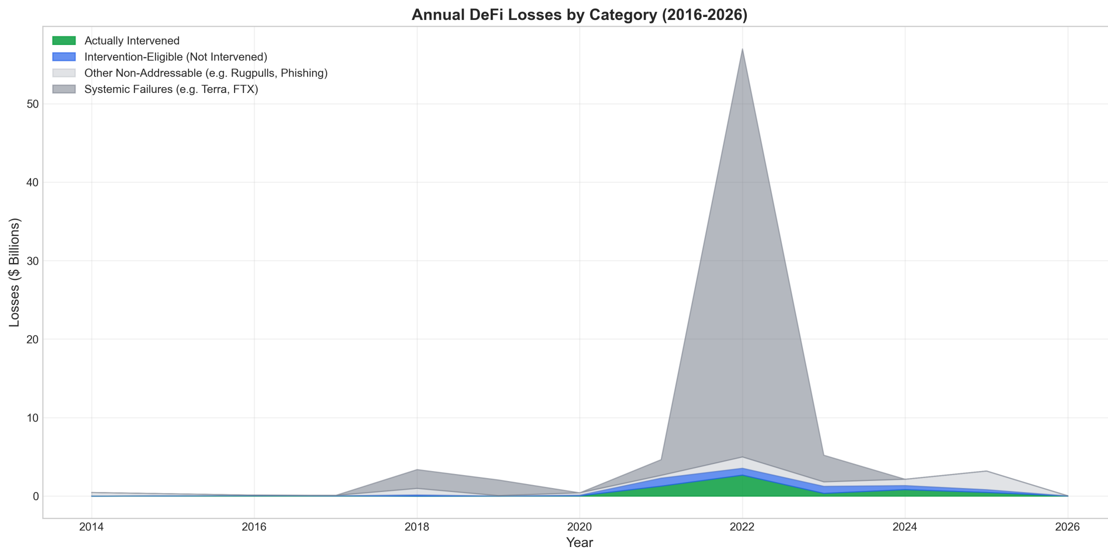

Figure 3: **Stratification of Losses (2016-2026).** We stratify losses into four layers: **Systemic Failures** (dark grey, e.g., Terra), **Other Non-Addressable** (light grey, e.g., rug pulls), **Intervention-Eligible** (blue), and **Actually Intervened** (green). This reveals that while systemic events dominate 2022, addressable technical exploits represent a consistent baseline of risk.

B Detailed Incident Analysis

We provide expanded narratives for representative intervention cases across the Scope × Authority taxonomy.

B.1 Network-scope interventions

Signer Set: BNB Chain and Harmony. During the October 2022 BNB Chain bridge exploit ("BSC Token Hub") [10, 35], validators coordinated to halt block production as an emergency containment measure. This intervention reduced further drainage but temporarily suspended unrelated activity on the network, making the collateral liveness cost immediately salient. Similar dynamics appeared in Harmony's Horizon Bridge exploit [30] and later Berachain's validator-coordinated halt [8].

Governance: The DAO and Gnosis Forks. Following the November 2025 Balancer exploit and subsequent fund freezes, Gnosis Chain executed a governance-approved hard fork (December 2025) to recover a reported $9.4M in assets that remained frozen onchain [29, 28].

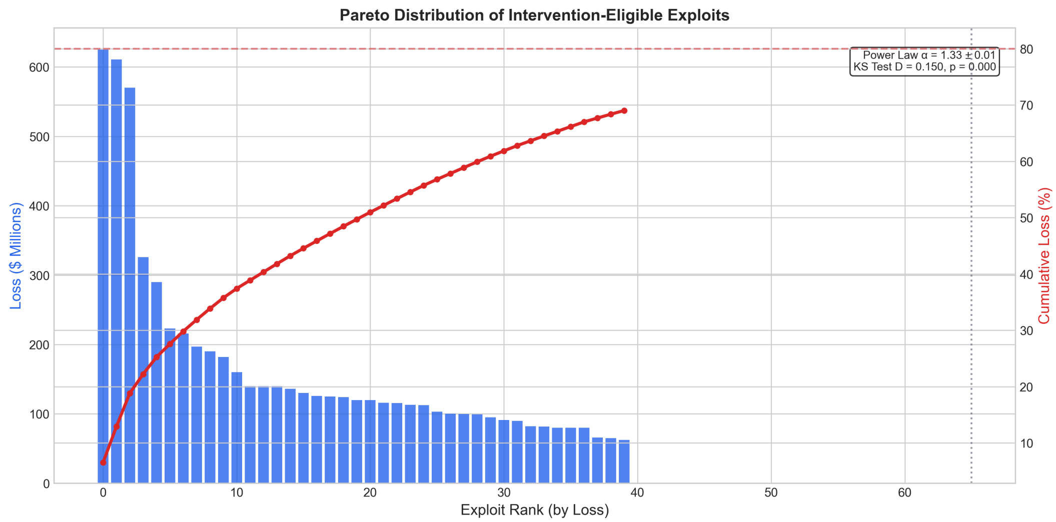

Figure 4: **Pareto Distribution of Intervention-Eligible Losses.** Approximately 80% of cumulative losses in our addressable dataset are attributable to fewer than 50 incidents. Power law fit: $\alpha \approx 1.33$, KS test $D = 0.150$, $p < 0.001$.

B.2 Asset-scope interventions

Signer Set: Tether and Circle. Circle's USDC terms explicitly describe "Blocked Addresses" and reserve the ability to freeze USDC [13]. In January 2026, Tether executed its largest-ever freeze, blocking $182M in USDT wallets [52].

Delegated Body: Bridge Governance. During the November 2025 Balancer exploit, the Gnosis Bridge Governance Board temporarily halted outflows of major tokens [28]. The Curve Emergency DAO provides another example with powers deliberately constrained to asset-specific actions [33].

B.3 Protocol-scope interventions

Signer Set: Admin Controls and Kill Switches. Liqwid's Proposal 44 (March 2024) explicitly granted the Core Team a single-signature "kill switch" to halt market batching within 5–15 minutes [32].

Delegated Body: Emergency subDAOs. Aave distinguishes governance from protocol guardians who hold time-bounded multisig authority [2, 3]. Curve Finance's July 2023 $62M exploit response was managed via its Emergency DAO [33]. Balancer V3 formalizes this further with explicit mandate limits [6].

Governance Process: ESM and Shutdown. MakerDAO's ESM mechanism has since been deprecated [42]. Euler Finance (March 2023) exemplified the "Protocol/Governance" tension: lacking an immediate admin override, recovery required offchain negotiation [22].

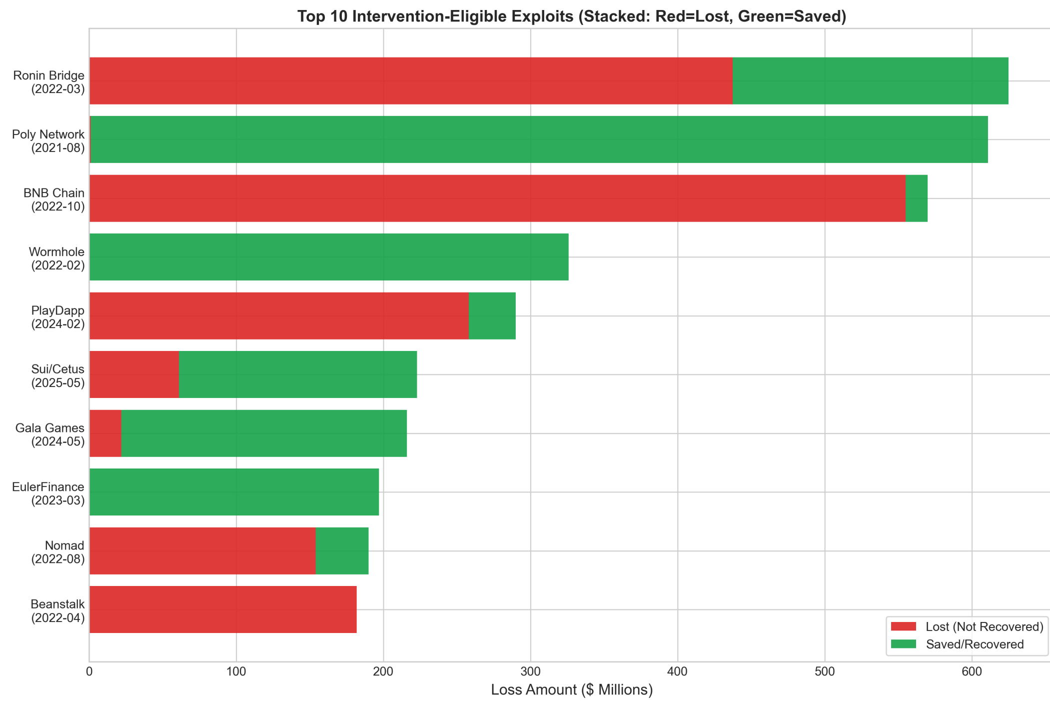

Figure 5: **Top 10 Intervention-Eligible Exploits.** Stacked bars show losses prevented (green) versus lost (red).

The Balancer Case and Window Expiry. The November 2025 Balancer V2 exploit highlights a core design trade-off: fixed-length pause windows improve immutability but reduce intervention capability [7, 6]. The Cork Protocol incident (May 2025) illustrates emergency war room coordination with SEAL911 [15].[2]

B.4 Module-scope interventions

Delegated Body: Feature-Specific Pauses. Aave's role-based control system supports pausing at the pool or reserve level [3, 1].

Governance Process: Function Toggles. MakerDAO's emergency governance proposal during the March 2023 USDC depeg [44] and dYdX's margin adjustments during the YFI incident [18] exemplify this pattern.

B.5 Account-scope interventions

Signer Set: Key-Based Targeted Restrictions. During the November 2025 Balancer exploit, Sonic Labs deployed `freezeAccount` within two hours [45]. However, 78.5% of funds were lost through a `permit()` bypass. After the March 2022 Ronin Bridge exploit ($625M), attacker addresses were partially frozen despite 6-day delayed detection [43].

[2]SEAL911 operates under the Whitehat Safe Harbor Agreement [40, 16, 37].

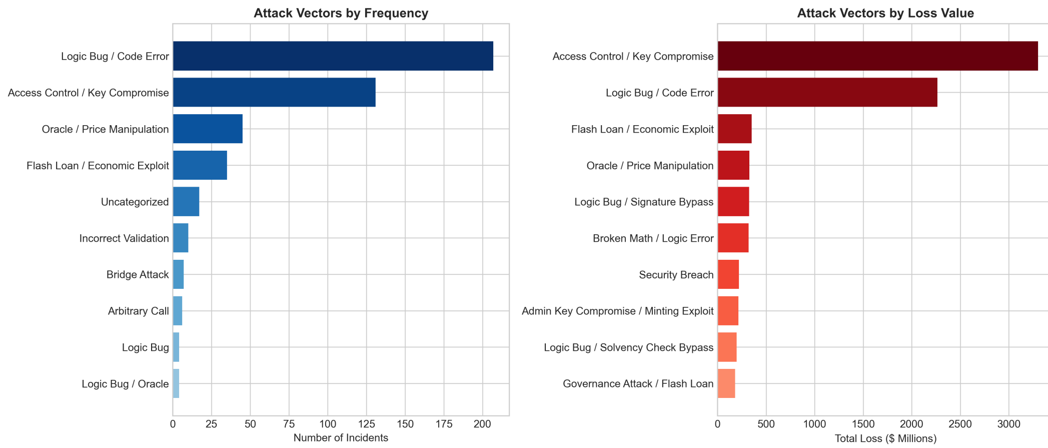

Figure 6: **Attack Vector Distribution.** While 'Logic Errors' and 'Access Control' issues are frequent and account for significant losses; complex 'Oracle Manipulation' and 'Flash Loan' attacks often result in the highest severity incidents.

Delegated Body: Council-Authorized Remediation. In the December 2025 Flow incident, an "Isolated Recovery" approach restricted only 1,060 addresses (under 0.01%) [26, 25]. The StakeWise emergency multisig recovered $20.7M without disrupting stable users [46]. Yearn's pxETH burn recovered $2.4M [53].

Governance Process: Stake-Weighted Vote. The May 2025 Cetus exploit on Sui produced the most rigorous governance-authorized account intervention: 90.9% of stake voted "Yes" [47, 48].

The VeChain Classification Dispute. VeChain's mechanism is a validator-enforced blocklist authorized by community governance, not a unilateral admin key [11, 49].

C Future Directions

- **Formalizing Heterogeneous Stakeholder Costs.** Different stakeholders bear asymmetric costs from both exploits and interventions.
- **Decision Support Tooling.** We have implemented an open-source prototype calculator for real-time mechanism calibration.
- **Extended Taxonomy Development.** Pre-execution prevention primitives suggest need for additional timing and enforcement dimensions.
- **Cross-Chain Coordination Protocols.** Standardized cross-chain emergency protocols could enable synchronized responses.
- **Dynamic Mechanism Selection.** Adaptive approaches based on real-time threat assessment.

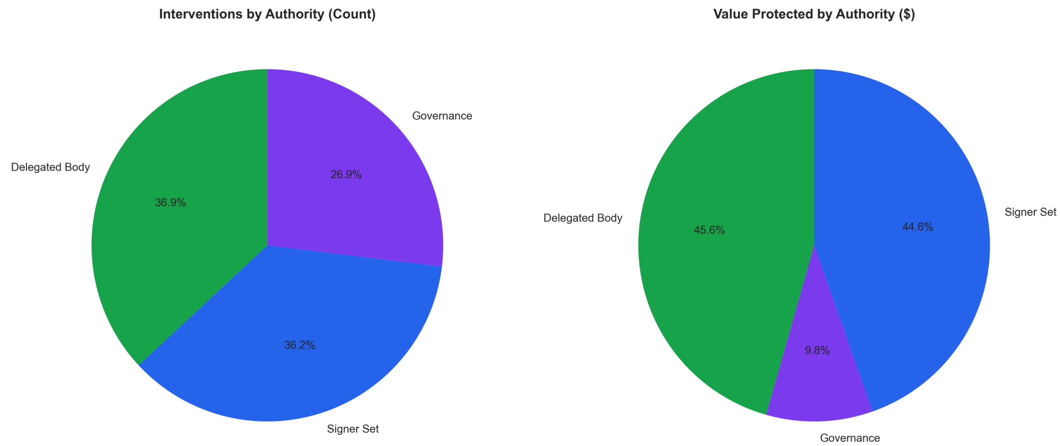

Figure 7: **Authority Distribution.** Signer Set dominates incident count, while Governance interventions achieve significant loss prevention through negotiation and recovery.

References

[1] Aave Governance. Full deprecation of dpi across aave deployments, 2025.

[2] Aave Protocol. Aave protocol guardians, 2023.

[3] Aave Protocol. Aave v3 acl manager: Role-based access control, 2023.

[4] Anchor Protocol Community. Emergency Measures for Restoring Terra Peg, May 2022. Emergency governance proposals during UST collapse.

[5] Lital Badash, Nachiket Tapas, Asaf Nadler, Francesco Longo, and Asaf Shabtai. Blockchain-based Bug Bounty Framework. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing (SAC '21)*, pages 239–248, New York, NY, USA, 2021. ACM.

[6] Balancer Foundation. Balancer v3 governance: Emergency subdao, 2024.

[7] Balancer Foundation. Nov 3 exploit post-mortem, Nov 2025. Detailed technical breakdown of CSP rounding bug, 94.8$Mtheft$,45.7M recovered/protected.

[8] Berachain Foundation. Berachain post-mortem: Validator halt and recovery, Nov 2025.

[9] Bit Novosti. Ethereum classic: keep censorship-resistant ethereum going, 2016. Ethereum Classic announcement following the DAO fork.

[10] BNB Chain. Bnb chain: A decentralized response, Oct 2022.

[11] ByBit Research. Blockchain security report 2024: Asset freezing landscape, 2025. Analysis of chain-level freezing capabilities.

Figure 8: **Scope × Authority Heatmap.** Intervention effectiveness (containment success %) across the taxonomy. Protocol-scope interventions are most frequent, while account-scope actions show high precision.

[12] Ben Charoenwong and Mario Bernardi. A decade of cryptocurrency "hacks": 2011–2021, 2021. SSRN working paper, revised 2025-11-02.

[13] Circle Internet Financial. Circle usdc terms of service: Blocked addresses, 2025.

[14] CoinDesk. World liberty financial blacklists justin sun's address with $107m wlfi, Sep 2025.

[15] Cork Protocol. Cork protocol post-mortem, May 2025.

[16] Dedaub. Seal 911: A few lessons from the frontlines, 2025. Accessed: February 2026.

[17] Kaustubh Dwivedi, Ankit Agrawal, Ashutosh Bhatia, and Kamlesh Tiwari. A Novel Classification of Attacks on Blockchain Layers: Vulnerabilities, Attacks, Mitigations, and Research Directions, 2024.

[18] dYdX. dydx: Sushi/yfi market incident report, Nov 2023. YFI market paused during large position liquidation. Module-level circuit breaker.

[19] Elliptic. The poly network hack: $600 million in crypto stolen and returned in 24 hours, Aug 2021.

[20] Ethereum Foundation. Critical update re: Dao vulnerability, 2016. The DAO was a smart contract on Ethereum that raised $150M and was exploited in June 2016, leading to the Ethereum hard fork.

[21] Ethereum Foundation. The dao hard fork (2016): Ethereum foundation blog, 2016.

[22] Euler Finance. War & Peace: Behind the Scenes of Euler's $240M Exploit Recovery, April 2023. No emergency pause; recovery via negotiation. Attacker returned $143M after negotiation.

[23] Euler Finance and Phylax Systems. Euler finance "holy grail" assertion: Account liquidity invariant, 2026. Euler deployed five assertions to protect lending protocol; primary assertion prevents healthy accounts from becoming liquidatable.

[24] John Ferejohn and Pasquale Pasquino. The law of the exception: A typology of emergency powers. *International Journal of Constitutional Law*, 2(2), 2004.

[25] Flow Foundation. Flow network recovery: Technical implementation plan, Dec 2025. Technical details of the Isolated Recovery plan.

[26] Flow Foundation. Flow security incident 27th december: Technical post-mortem, Jan 2026.

[27] Ilana Gimpelson, G Karavokkyris, I Lachman, G Lurie, M Pacholska, T Shwartz, Y Orpeli, A Reichman, E Salzberger, G Barzilai, et al. Law and emergencies: A comparative overview, the minerva center for the rule of law under extreme conditions. *Available online also at: http://minervaextremelaw. haifa. ac. il/images/Emergency_Laws_and_Regulations_-in_Japan-_19-_Jan2016. pdf [accessed in Bandung, Indonesia: March 2, 2016]*, 2016.

[28] Gnosis Bridge Governance Board. Bridge board decision: Freeze outflow of major tokens on canonical bridges, Nov 2025. Due to BalancerV2 exploit, halted outflow of major tokens on Omnibridge & xDAI bridge as precautionary measure.

[29] GnosisDAO. Balancer hack hard fork proposal, Dec 2025.

[30] Halborn. Explained: The Harmony Horizon Bridge Hack (June 2022), June 2022. Horizon bridge compromised via key theft. $100M loss.

[31] Clayton Hutto and Eric Gilbert. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8, pages 216–225, 2014.

[32] Liqwid Finance Governance. Proposal 44: Grant core team power to rapidly deal with emergencies, March 2024. Proposal 44, Executed March 14, 2024.

[33] LlamaRisk. Curve finance exploit analysis: Emergency subdao response, Jul 2023. Emergency DAO freeze CRV gauge emissions to affected pools.

[34] Junjie Ma, Muhui Jiang, Jinan Jiang, Xiapu Luo, Yufeng Hu, and Yajin Zhou. Understanding Security Issues in the DAO Governance Process. *IEEE Transactions on Software Engineering*, 51(4):1188–1204, April 2025. Analysis of 3,348 DAOs across 9 blockchains revealing governance contract backdoors and malicious proposals.

[35] Merkle Science. Hack track: Analysis of the bnb smart chain exploit, Oct 2022.

[36] Phylax Systems. Phylax partners with linea: Bringing network-native security to the home of eth capital, Jan 2026. Linea integrates Credible Layer for pre-execution exploit prevention via sequencer-enforced assertions.

[37] Piper Alderman. Security alliance proposes whitehat safe harbor to secure web3, 2025. Accessed: February 2026.

[38] Lori Qian. Strengthening DAO Governance: Vulnerabilities and Solutions. *The National High School Journal of Science*, 2025. Case study analysis of Uniswap, GnosisDAO, and ArbitrumDAO governance vulnerabilities including flash loan exploitation and off-chain voting manipulation.

[39] Reuters. Binance-linked blockchain hit by $570 million crypto hack, Oct 2022.

[40] Security Alliance. Seal-911 github repository, 2024. Accessed: February 2026.

[41] Md Kamrul Siam, Bilash Saha, Md Mehedi Hasan, Md Jobair Hossain Faruk, Nafisa Anjum, Sharaban Tahora, Aiasha Siddika, and Hossain Shahriar. Securing Decentralized Ecosystems: A Comprehensive Systematic Review of Blockchain Vulnerabilities, Attacks, and Countermeasures and Mitigation Strategies. *Future Internet*, 17(4):183, 2025.

[42] Sky / MakerDAO. Emergency shutdown (deprecated): Sky protocol documentation, 2025. Official documentation for the deprecated Emergency Shutdown Module (ESM).

[43] Sky Mavis. Back to Building: Ronin Security Breach Post-Mortem, March 2022. $625M bridge exploit. $30M recovered via law enforcement.

[44] Sky Money Forum. Emergency proposal: Risk and governance parameter changes (march 11, 2023), Mar 2023. Emergency governance vote to pause PSM during USDC depeg.

[45] Sonic Labs. Sonic chain: Defensive measures and the freeze post-mortem, Nov 2025.

[46] StakeWise DAO. Post-mortem: Stakewise dao recovery of $20.7m ostokens from balancer v2 exploiter, Nov 2025. Emergency multisig (7 members) used token controller roles to burn hijacked osETH/osGNO and re-mint to DAO addresses.

[47] Sui Foundation. Response to the cetus incident - onchain community vote, May 2025.

[48] Unchained Crypto. Cetus relaunches protocol after recovering $162m from exploit, Jun 2025.

[49] VeChain Foundation. Vechain refutes bybit's allegations: Clarifying freezing vs blocking, 2025. Official clarification on freezing capabilities.

[50] Qin Wang, Guangsheng Yu, Yilin Sai, Caijun Sun, Lam Duc Nguyen, and Shiping Chen. Understanding DAOs: An Empirical Study on Governance Dynamics. *IEEE Transactions on Computational Social Systems*, 12(5):2814–2832, October 2025. Empirical analysis of 581 DAO projects and 16,246 proposals from Snapshot.

[51] Kevin Werbach, Primavera De Filippi, Joshua Tan, and Gina Pieters. Blockchain Governance in the Wild. *Cryptoeconomic Systems*, April 2024. Comparative questionnaire study of governance practices at 23 blockchain projects examining on-chain and off-chain mechanisms.

[52] Yahoo Finance. Tether freezes $182m usdt in largest-ever freeze, Jan 2026. Tether froze wallets linked to Venezuelan state oil company PDVSA.

[53] Yearn Finance Security Team. Yearn finance post-mortem: yeth exploit and recovery, Nov 2025.