

Blockchain-Anchored Detection and Recovery for Resilient Reinforcement Learning-Based Cyber-Physical Systems

Imran Jahid Khan

Department of Computer Science, Tuskegee University

Maliha Hossain

Department of Computer Science, Tuskegee University

Mohammad Rahman

Department of Computer Science, Tuskegee University

Fan Wu

Department of Computer Science, Tuskegee University

Author Note

Correspondence concerning this article should be addressed to Imran Jahid Khan, Department of Computer Science, Tuskegee University, Tuskegee, Alabama, USA.

Email: mkhan7037@tuskegee.edu

Imran Jahid Khan: mkhan7037@tuskegee.edu

Maliha Hossain: mhossain8750@tuskegee.edu

Mohammad Rahman: mrahman@tuskegee.edu

Fan Wu: fwu@tuskegee.edu

<https://doi.org/10.68414/OMVN2247>

Abstract

Reinforcement learning (RL) controllers in cyber-physical systems (CPS) assume that sensor observations describe the true plant state, yet replay, sensor-drift, and false-data-injection (FDI) attacks can corrupt the observation stream at inference time, and detection alone does not tell the controller what to use after an alarm. This study evaluates a proof-of-concept defense that combines a SHA-256 hash-linked ledger, which provides exact-match replay detection and a tamper-evident recovery anchor, with a Long Short-Term Memory autoencoder (LSTM-AE) for audit alarms and a fixed Proximal Policy Optimization (PPO) controller. Ten independently trained policies per environment are evaluated in paired episodes on CartPole-v1 and LunarLander-v3. Content-hash matching detects every injected bitwise replay with zero false matches in 379,396 clean steps. However, perfect detection does not guarantee safe recovery: exact-gated ledger hold raises LunarLander replay return from 178.4 to 264.6 (paired mean +86.2, 95% CI [64.8, 107.6]) but lowers CartPole from 411.1 to 376.2 (−34.9, not significant after Holm adjustment). A trigger-by-anchor factorial ablation shows that local and ledger

anchors are control-equivalent under a fixed trigger, while LSTM-AE-gated hold is harmful in every tested cell. A 4,800-episode sensitivity study across three severity levels per attack shows that the recovery effect scales with severity in opposite directions: +5.8 to +225.0 on LunarLander and −4.6 to −60.3 on CartPole. The ledger contributes exact-replay lookup, provenance, and tamper evidence rather than an automatic control advantage; safe general recovery further requires dynamics-aware substitution and distributed, physical validation.

Keywords: cyber-physical systems, reinforcement learning, blockchain, anomaly detection, replay attacks, resilient control

Introduction

RL-controlled CPS couple sensing, computation, and actuation: a policy maps observation $\mathbf{o}_t$ to action a_t , and the physical process generates the next state. If an attacker manipulates the observation channel, a competent clean policy may act on a false description of the plant. Replay substitutes a historical packet, drift accumulates a small bias, and false data injection (FDI) fabricates selected features. Detection alone is insufficient because the controller still needs a usable input after an alarm; conversely, a recovery state without verifiable provenance may already be compromised.

This paper evaluates a narrow integration of three functions. A hash-linked ledger performs exact content lookup and records detector flags; an LSTM-AE supplies temporal audit alarms; and an exact-match trigger retrieves the latest earlier eligible ledger observation before PPO inference. The implementation is intentionally a single-node prototype, not a production blockchain. Its evidence concerns the data path—exact matching, hash-linked tamper evidence, and recovery-state provenance—not distributed consensus or Byzantine fault tolerance.

This study makes four contributions. First, step-level ground truth demonstrates zero false positives and false negatives for the evaluated bitwise replay, while explicitly excluding noisy or semantically equivalent replay. Second, paired policy-level inference shows that the same recovery improves LunarLander but can harm CartPole; perfect detection is therefore not sufficient for safe recovery. Third, a full trigger (exact or learned) by anchor (local or ledger) ablation separates control effect from provenance. Fourth, an attack-severity sensitivity study adds 4,800 episodes across three attack levels and shows that recovery utility varies with attack severity and plant dynamics.

Literature Review

Adversarial CPS research identifies the observation interface as a major attack surface. Li et al. (2020) organize attacks and defenses across heterogeneous sensor, communication, and computational interfaces, showing that CPS exposure extends beyond conventional image classifiers. Empirical studies reinforce this concern. Figueroa et al. (2022) evaluate FGSM, DeepFool, and JSMA perturbations against machine-learning models used in industrial-control CPS and report substantial losses in predictive reliability. Gipiškis et al. (2023) similarly show that adversarial perturbations can alter both semantic-segmentation outputs and their saliency-based explanations. Together, these studies establish that a controller may receive a plausible but operationally misleading observation, motivating defenses that address the observation delivered to the control policy.

Anomaly-detection research offers broad coverage but does not by itself resolve what the controller should use after an alarm. In a review of 296 CPS studies, Jeffrey et al. (2023) classify defenses spanning intrusion detection and prevention, machine learning, simulation and testbeds, edge processing, and zero-trust architectures. Jeffrey et al. (2024) combine signature, threshold, and machine-learning detectors and find one-class KNN effective for imbalanced CPS data. Such approaches strengthen recognition of abnormal behavior, but their evaluation does not isolate state substitution inside a closed-loop RL controller. The distinction is important because a correct alarm can still be followed by a stale or dynamically unsafe replacement observation.

Recovery research addresses that post-detection problem more directly. L. Zhang et al. (2021) combine constrained LQR control, state reconstruction, checkpointing, and deadline estimation, demonstrating recovery from bias, replay, and delay attacks across five CPS simulators. Their results illustrate that recovery requires both a trustworthy state estimate and a control action that respects system dynamics. Reviewing work published from 2014 to 2023, Lu et al. (2024) conclude that CPS recovery remains less mature than detection and is concentrated on sensor and communication attacks. Controller-, actuator-, and application-specific recovery therefore remain open concerns. The present study examines a narrower question: whether a previously verified observation can support inference-time recovery for an already trained RL controller.

Blockchain-based CPS frameworks primarily address authentication, confidentiality, integrity, and accountability. Rahman et al. (2021) combine consortium blockchain,

multi-signature authentication, edge computing, and distributed storage for Industry 4.0 data exchange. Ali et al. (2023) integrate smart contracts, access control, cryptographic protection, and decentralized validation for privacy-sensitive CPS information, while Bhattacharjya (2022) synthesizes blockchain’s role in preserving confidentiality, integrity, and availability across CPS and IoT communication. These studies motivate a ledger as a source of provenance, but they do not separate the control effect of state retrieval from the trust benefit of the ledger itself.

Robust-RL research provides a complementary defense strategy. H. Zhang et al. (2021) show that small state-observation perturbations can sharply reduce policy return and propose alternating training with learned adversaries to reduce that sensitivity. Chen and Li (2020) survey robust RL under transition-model uncertainty and emphasize the difficulty of deploying learned policies when estimated dynamics differ from the operating environment. These methods modify training to improve policy robustness. In contrast, the present work keeps the clean PPO policy fixed and studies inference-time detection, provenance, and substitution, so the two approaches address different points in the defense lifecycle.

The resulting gap is not a lack of individual detectors, recovery algorithms, or blockchain mechanisms, but limited evidence connecting all three to closed-loop RL behavior. An integrated evaluation must distinguish detection accuracy, recovery-state provenance, the dynamics of the substitution rule, and the controller’s resulting return. Table 1 positions the proposed proof of concept against closely related CPS security frameworks and highlights its trigger-by-anchor ablation, which separates the recovery policy from the storage location.

Table 1

Comparison with Closely Related Integrated CPS Frameworks

Study	Principal mechanism	Detection	Recovery	Ledger	RL / ablation
Jeffrey et al. (2024)	Hybrid IDS	Yes	No	No	No / No
Zhang et al. (2021)	LQR recovery	Yes	Yes	No	No / No
Rahman et al. (2021)	Consortium authentication	No	No	Yes	No / No
Ali et al. (2023)	Consortium privacy	No	No	Yes	No / No
Proposed	Exact/AE + hold	Yes	Yes	Yes	Yes / Yes

Note. “Ledger” denotes a blockchain or hash-linked record used to establish data origin/integrity. The last column reports RL closed-loop evaluation / trigger–anchor ablation.

Methodology

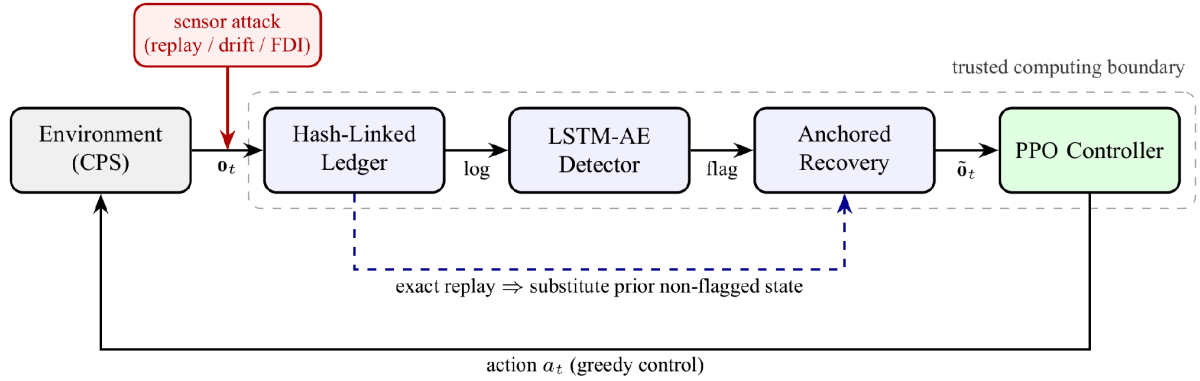
Threat Model and Scope

The adversary can alter the sensor packet delivered to the controller but cannot change plant dynamics, PPO parameters, ledger code, or the simulator-private clean state. Replay draws a bitwise float32 copy from previously *delivered* packets delayed by 20 steps; this avoids giving the attacker privileged clean-state access. Drift and FDI affect continuous dimensions only; LunarLander contact bits remain unchanged. The ledger and controller are inside the trusted computing boundary. Network compromise, key theft, actuator attacks, collusion among permissioned validators, and database-administrator rewrite are outside the experiment.

Figure 1 summarizes the implemented data path and experimental trust boundary.

Figure 1

System Architecture and Experimental Trust Boundary



Note. Per-timestep data flow. The LSTM-AE flag is committed for audit but does not trigger the primary substitution path; exact replay retrieves the latest earlier observation eligible under the exact trigger.

Ledger, Detection, and Recovery

Each block stores the observation; node, session, episode, and timestep identifiers; replay and AE flags; the prior block hash; a current SHA-256 hash over state, metadata, prior hash, and flags; and a content hash over observation values and node identity. The content hash excludes time so a bitwise repeated packet matches an earlier record. Replay lookup occurs before the current block is committed. Session-aware uniqueness constraints prevent duplicate timesteps, and the chain is recomputed after every episode.

The hash chain is tamper-evident, not immutable. A single administrator who rewrites all records and recomputes hashes can defeat the prototype. Distributed deployment would require

authenticated membership, replicated storage, consensus such as Raft or PBFT, and independently held checkpoints (Castro & Liskov, 1999; Ongaro & Ousterhout, 2014).

The detector is a two-layer sequence-to-sequence LSTM-AE (hidden size 64; window $W = 20$). With training-standardized observation $\mathbf{z}$, its score is

$$s_t = \frac{1}{Wd} \sum_{i=1}^W \|\mathbf{z}_{t-W+i} - \hat{\mathbf{z}}_i\|_2^2. \quad (1)$$

For each environment, clean trajectories from all deployed policy seeds are split by whole episode before overlapping windows are formed, preventing adjacent-window leakage. Training/validation contain 300/100 episodes and 144,300/48,100 CartPole windows or 68,634/23,921 LunarLander windows. The threshold is the held-out 99.9th percentile (CartPole .01649; LunarLander 1.54518), and an audit alarm requires 20 consecutive exceedances. Because one corrupted sample contaminates up to 20 overlapping windows, this persistence gate can remain active after an intermittent event and is not used for primary recovery.

On an exact match, the controller receives the latest earlier block whose replay flag is zero:

$$\tilde{\mathbf{o}}_t = \begin{cases} \mathbf{o}_t, & \text{if no exact replay is detected,} \\ \mathbf{o}_t^{\text{anchor}}, & \text{if an exact replay is detected.} \end{cases} \quad (2)$$

Anchor eligibility is trigger-specific: an audit-only AE flag cannot silently remove an anchor from the exact path. “Eligible” is operational and does not prove that the physical state is clean.

Controller and Experimental Design

The PPO controller (Schulman et al., 2017) uses a shared two-layer, 128-unit tanh trunk and separate policy/value heads; learning rate decreases linearly from 3×10^{-4} , with $\gamma = .99$, GAE $\lambda = .95$, clip .20, ten epochs, and batch size 64. Observations are standardized online during clean training and frozen for evaluation. Time-limit truncations are bootstrapped, current-transition termination masks GAE, and partial rollouts are optimized. Ten independent policies (seeds 1000–1009) are trained per environment. In a disjoint 20-episode clean validation, all CartPole policies score 500 and every LunarLander policy exceeds 200 mean return.

Table 2 summarizes the task, replication, and attack-severity settings.

Table 2*Primary Experiment and Sensitivity Settings*

Component	CartPole-v1	LunarLander-v3
Observation dimensions / discrete actions	4 / 2	8 / 4
Clean mean return	500.0	265.7
Independent policies	10	10
Primary episodes per policy/condition	50	50
Sensitivity episodes per policy/cell	20	20
Replay probability (low/base/high)	.10 / .30 / .50	.10 / .30 / .50
Drift rate, feature scales/step	.001 / .002 / .004	.001 / .002 / .004
FDI <i>SD</i> , feature scales	1.5 / 3.0 / 6.0	1.5 / 3.0 / 6.0

Note. Replay uses a 20-step delay. FDI is attempted with probability .20 and applied independently to selected continuous dimensions with probability .50.

The primary study evaluates clean baseline, attack without defense, and attack with exact-gated ledger recovery. Each policy receives 50 episodes per condition. Environment seeds and independent attack-schedule/payload seeds are reused across defense conditions. A factorial ablation crosses recovery trigger (exact hash or persistence-gated AE) with anchor source (local cache or ledger), plus a no-recovery reference. The sensitivity study holds every non-varied attack parameter at baseline and adds 20 paired episodes per policy and severity cell: 4,800 episodes total.

The independently trained policy is the inferential unit. For each of six primary environment–attack contrasts, we report the paired mean difference, 95% paired t CI, $t(9)$, exact two-sided enumeration of all 2^{10} sign flips, paired d_z , a 10,000-resample hierarchical bootstrap CI (policies, then episodes within policy), and Holm-adjusted t -test p values. Descriptive variability is the *SD* of ten policy means. Sensitivity replay contrasts use the same policy-level paired analysis and Holm correction across its six cells.

Results and Findings**Main Effects and Ablation**

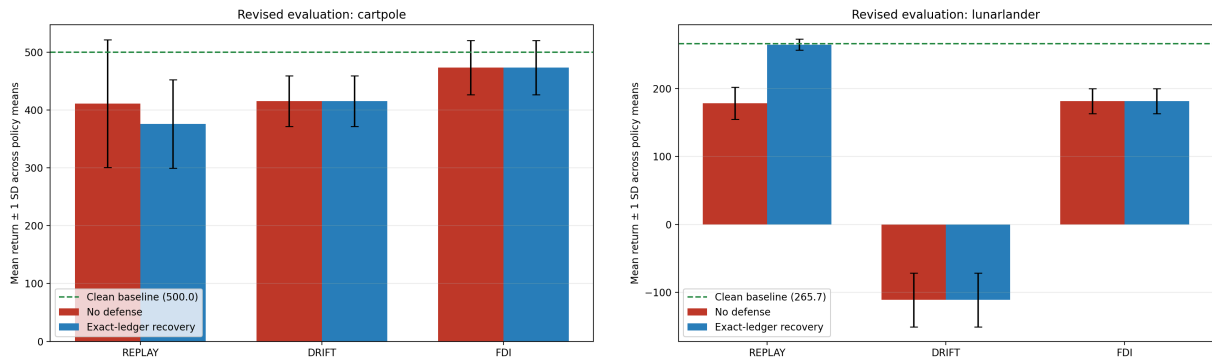
Exact-gated recovery improves LunarLander replay from 178.4 to 264.6. All ten policy differences are positive; the paired difference is +86.2, 95% CI [64.8, 107.6], $t(9) = 9.11$, $p < .001$, exact sign-flip $p = .001953$, $d_z = 2.88$, hierarchical bootstrap CI [64.0, 106.8], and Holm-adjusted $p < .001$. This restores 99.6% of the clean mean.

CartPole changes from 411.1 to 376.2: -34.9 , CI $[-65.6, -4.2]$, $t(9) = -2.57$, $p = .030$, exact $p = .035$, $d_z = -0.81$, bootstrap CI $[-63.1, -6.2]$. Seven of ten differences are negative, but Holm-adjusted $p = .150$ does not reject across six primary contrasts. Drift and FDI returns are identical between primary defended and undefended conditions because exact-only recovery correctly abstains; this is scope control, not non-replay recovery.

Figure 2 visualizes the primary evaluation returns across attacks and environments.

Figure 2

Primary Evaluation Return by Attack and Environment



Note. CartPole left, LunarLander right. Bars are means of ten policy means; error bars are ± 1 SD. The dashed line marks clean mean return. Exact-only recovery abstains on drift and FDI, so defended and undefended bars coincide in those cells.

Table 3 reports the trigger-by-anchor ablation across both environments and all three attacks.

Table 3

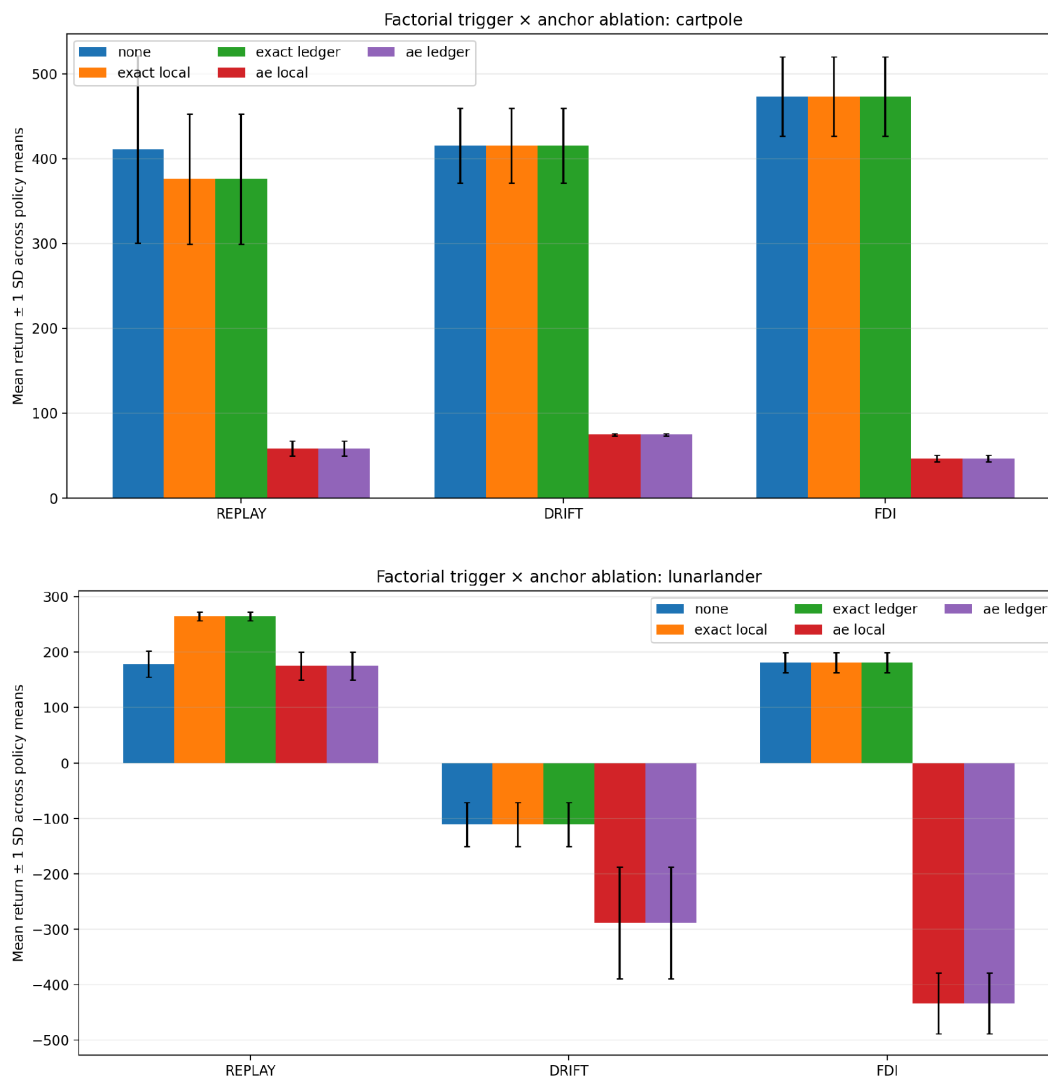
Trigger-by-Anchor Factorial Ablation: Mean Return $\pm$ SD Across Ten Policy Means

Environment	Attack	No recovery	Exact: local = ledger	AE: local = ledger
CartPole	Replay	411.1 $\pm$ 110.4	376.2 $\pm$ 76.5	58.3 $\pm$ 9.1
	Drift	415.4 $\pm$ 44.0	415.4 $\pm$ 44.0	74.6 $\pm$ 1.6
	FDI	473.2 $\pm$ 46.8	473.2 $\pm$ 46.8	46.8 $\pm$ 3.6
LunarLander	Replay	178.4 $\pm$ 23.7	264.6 $\pm$ 8.2	175.0 $\pm$ 24.8
	Drift	-111.2 $\pm$ 39.8	-111.2 $\pm$ 39.8	-288.4 $\pm$ 100.9
	FDI	181.2 $\pm$ 18.4	181.2 $\pm$ 18.4	-433.7 $\pm$ 54.9

Note. Local and ledger anchors return the same newest eligible packet under a fixed trigger; the maximum absolute paired policy-mean difference is 0 in all six cells.

Anchor location alone never changes return. The ledger’s independent contribution is provenance and tamper evidence. Trigger choice is decisive: AE-gated hold harms all tested cells because overlapping-window persistence can substitute a stale state after the corrupt event. Exact replay recovery is active on 28.4% of CartPole and 27.5% of LunarLander steps; mean anchor age is only 1.49 and 1.42 steps. Thus, even a recent zero-order hold can be unsafe in faster dynamics. The exact equality of the local and ledger columns is expected rather than anomalous: under a fixed trigger, both anchors deliver the bitwise-identical newest eligible packet, so the closed-loop trajectory cannot differ. Storage location determines who can verify and audit the anchor, not the numeric value the controller receives; the ledger’s tamper evidence therefore matters for trust and forensics rather than for reward. The environment asymmetry also has a mechanical reading. At a .30 replay probability, roughly one in four controller inputs is a held state, and CartPole’s short, unstable balancing horizon amplifies even one-step-old substitutions into accumulating corrective errors. LunarLander’s slower descent both tolerates a recent hold and benefits from suppressing packets that are 20 steps stale, which is why the same recovery rule moves the two environments in opposite directions.

Figure 3 displays the complete factorial ablation for both control tasks.

Figure 3*Trigger-by-Anchor Factorial Ablation by Environment*

Note. CartPole top, LunarLander bottom. Bars are means of ten policy means; error bars are ± 1 *SD*. Exact-local and exact-ledger bars coincide, as do AE-local and AE-ledger bars, because anchor eligibility is held fixed within each trigger.

Detection and Prototype Timing

Exact matching detects all 85,229 injected replay packets with zero false positives or false negatives and zero matches in 379,396 clean steps; first-event delay is zero. The guarantee applies only to the evaluated bitwise representation. The AE gate has zero false positives on clean evaluation but poor temporal localization under intermittent attacks: CartPole replay precision/recall/FPR are .300/.914/.846, whereas LunarLander replay recall is below .001. For

FDI, FPR is .856 on CartPole and .793 on LunarLander because earlier corruptions remain inside the window. These values explain why the AE remains audit-only.

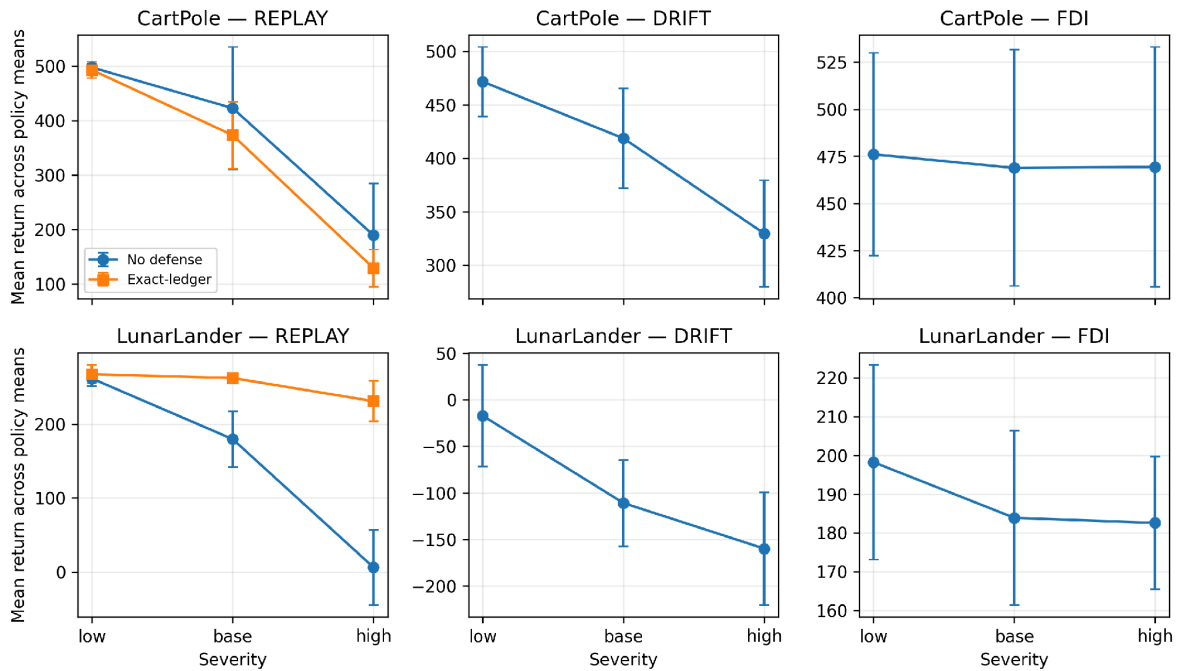
All episode chains verify. In defended replay, mean block-commit/query/anchor-retrieval times are .083/.024/.009 ms on CartPole and .090/.027/.009 ms on LunarLander; corresponding mean episode 95th percentiles are .138/.041/.028 and .148/.044/.029 ms. These in-memory SQLite timings exclude network transport, durable replication, consensus, and sensor/actuator delay.

Attack-Parameter Sensitivity

Figure 4 summarizes policy-level return across the low, baseline, and high attack levels.

Figure 4

Return Across Low, Baseline, and High Attack Severity



Note. Points are means of ten policy means; error bars are ± 1 SD. Replay includes both no-defense and exact-ledger conditions. Drift and FDI are audit-only in the primary design, so only no-defense curves are shown.

Replay sensitivity confirms a dose-response interaction rather than task-independent improvement. On LunarLander, the defended-minus-undefended effect is +5.8 at $p = .10$ (CI $[-3.1, 14.7]$, exact $p = .174$), +82.7 at $p = .30$ (CI $[52.9, 112.5]$, exact $p = .001953$), and +225.0 at $p = .50$ (CI $[183.2, 266.9]$, exact $p = .001953$). On CartPole it is -4.6 (CI $[-11.6, 2.4]$, exact p

= .500), -49.6 (CI $[-92.9, -6.2]$, exact $p = .039$), and -60.3 (CI $[-108.0, -12.6]$, exact $p = .023$), respectively. Exact matching retains zero false positives and negatives at every severity.

Undefended drift shows monotonic degradation: CartPole means are 471.7, 418.7, and 329.7; LunarLander means are -17.1 , -110.9 , and -159.8 . FDI is comparatively flat over the tested standard deviations: 476.1/468.9/469.4 on CartPole and 198.2/183.9/182.6 on LunarLander. The sweep strengthens parameter robustness but does not establish generalization to other CPS tasks or physical noise models.

Discussion

The combined evidence separates three claims often conflated in integrated security systems. First, content hashing is a deterministic detector for exact packet reuse under a stable serialization; it is not a detector for approximate replay. Second, a hash-linked anchor provides verifiable provenance but does not guarantee that its observation is physically correct. Third, recovery quality is a closed-loop property. LunarLander benefits as replay becomes more frequent, whereas CartPole is increasingly harmed by substituting recent historical states. Detector accuracy alone cannot decide whether substitution is safe.

The factorial experiment also prevents attributing the LunarLander gain to “blockchain” in the abstract. Local and ledger anchors are numerically identical under fixed eligibility. The ledger adds an auditable, tamper-evident history and exact lookup; the hold policy produces the control effect. Conversely, the AE negative result is useful: a high clean-data threshold and persistence gate can still localize intermittent events poorly. Non-replay recovery needs an event-localized trigger and a dynamics-aware estimate with uncertainty-based abstention.

Implications, Limitations, and Future Research

Seven limitations bound the conclusions. (1) Recovery handles exact bitwise replay only. (2) Zero-order hold is not dynamics-aware and harms CartPole. (3) The severity sweep varies one parameter at a time and does not cover adaptive or multivariate attacks. (4) The AE is calibrated on the evaluated policy family; transfer after policy or plant shift is unknown. (5) Ten policies in two simulators support run-level replication but not broad CPS generalization. (6) A single-node ledger provides tamper evidence, not distributed immutability. (7) Latency excludes networking, replication, consensus, and hardware. Next work should evaluate approximate-replay signatures, model-based state prediction with calibrated uncertainty, joint

attack sweeps, independent validators, additional control plants, and hardware-in-the-loop timing.

Conclusion

This paper evaluates exact replay detection, verifiable recovery anchoring, and RL control in one reproducible data path. Across two environments and ten policies each, exact matching detects every evaluated bitwise replay without false matches. Exact-gated hold improves LunarLander but can harm CartPole, and its effect grows in opposite directions as replay probability increases. The ablation shows that anchor location does not change return when eligibility is fixed; the ledger’s distinct value is lookup, provenance, and tamper evidence. The appropriate conclusion is therefore narrow but technically useful: exact replay is detectable, ledger-anchored hold can restore one tested control task, and safe general recovery requires dynamics-aware substitution plus distributed and physical validation.

Acknowledgments

This work was partially supported by the National Science Foundation under Awards 2019561, 2234911, 2209637, and 2100134. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

During preparation of the camera-ready manuscript, the authors used OpenAI ChatGPT and OpenAI Codex for grammar, spelling, style, and sentence-level clarity improvements throughout the manuscript; assistance with LaTeX formatting; and the generation and refinement of TikZ code used to produce Figure 1. The research design, methodology, implementation, experiments, data analysis, results, interpretations, and conclusions were produced by the authors. All AI-assisted text, formatting, and code were reviewed, revised, and verified by the authors, who take full responsibility for the manuscript.

References

- Ali, A., Al-rimy, B. A. S., Almazroi, A. A., Alsubaei, F. S., Almazroi, A. A., & Saeed, F. (2023). Securing secrets in cyber-physical systems: A cutting-edge privacy approach with consortium blockchain. *Sensors*, 23(16), 7162. <https://doi.org/10.3390/s23167162>
- Bhattacharjya, A. (2022). A holistic study on the use of blockchain technology in CPS and IoT architectures maintaining the CIA triad in data communication. *International Journal of Applied Mathematics and Computer Science*, 32(3), 403–413. <https://doi.org/10.34768/amcs-2022-0029>
- Castro, M., & Liskov, B. (1999). Practical byzantine fault tolerance. *Proceedings of the 3rd Symposium on Operating Systems Design and Implementation*, 173–186. <https://www.usenix.org/conference/osdi-99/practical-byzantine-fault-tolerance>
- Chen, S., & Li, Y. (2020). An overview of robust reinforcement learning. *2020 IEEE International Conference on Networking, Sensing and Control*, 1–6. <https://doi.org/10.1109/ICNSC48988.2020.9238129>
- Figuerola, H., Wang, Y., & Giakos, G. C. (2022). Adversarial attacks in industrial control cyber physical systems. *2022 IEEE International Conference on Imaging Systems and Techniques*, 1–6. <https://doi.org/10.1109/IST55454.2022.9827763>
- Gipiškis, R., Chiaro, D., Preziosi, M., Prezioso, E., & Piccialli, F. (2023). The impact of adversarial attacks on interpretable semantic segmentation in cyber–physical systems. *IEEE Systems Journal*, 17(4), 5327–5334. <https://doi.org/10.1109/JSYST.2023.3281079>
- Jeffrey, N., Tan, Q., & Villar, J. R. (2023). A review of anomaly detection strategies to detect threats to cyber-physical systems. *Electronics*, 12(15), 3283. <https://doi.org/10.3390/electronics12153283>

- Jeffrey, N., Tan, Q., & Villar, J. R. (2024). A hybrid methodology for anomaly detection in cyber–physical systems. *Neurocomputing*, 568, 127068.
<https://doi.org/10.1016/j.neucom.2023.127068>
- Li, J., Liu, Y., Chen, T., Xiao, Z., Li, Z., & Wang, J. (2020). Adversarial attacks and defenses on cyber–physical systems: A survey. *IEEE Internet of Things Journal*, 7(6), 5103–5115.
<https://doi.org/10.1109/JIOT.2020.2975654>
- Lu, P., Zhang, L., Liu, M., Sridhar, K., Sokolsky, O., Kong, F., & Lee, I. (2024). Recovery from adversarial attacks in cyber-physical systems: Shallow, deep, and exploratory works. *ACM Computing Surveys*, 56(8), 211:1–211:31. <https://doi.org/10.1145/3653974>
- Ongaro, D., & Ousterhout, J. (2014). In search of an understandable consensus algorithm. *2014 USENIX Annual Technical Conference*, 305–319.
<https://www.usenix.org/conference/atc14/technical-sessions/presentation/ongaro>
- Rahman, Z., Khalil, I., Yi, X., & Atiquzzaman, M. (2021). Blockchain-based security framework for critical industry 4.0 cyber-physical system. *arXiv*.
<https://doi.org/10.48550/arXiv.2106.13339>
- Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv*. <https://doi.org/10.48550/arXiv.1707.06347>
- Zhang, H., Chen, H., Boning, D., & Hsieh, C.-J. (2021). Robust reinforcement learning on state observations with learned optimal adversary. *International Conference on Learning Representations*. <https://openreview.net/forum?id=sCZbhBvqQaU>
- Zhang, L., Lu, P., Kong, F., Chen, X., Sokolsky, O., & Lee, I. (2021). Real-time attack-recovery for cyber-physical systems using linear-quadratic regulator. *ACM Transactions on Embedded Computing Systems*, 20(5s), 79:1–79:24. <https://doi.org/10.1145/3477010>