# Situ-HMARL: A Situation-Based Hierarchical Multi-Agent Reinforcement Learning Adaptive Defense Framework for Industrial Internet of Things Security

Xingke Zhu[1], Zhiyong Zhang[1,*], Haonan Wang[1], Bin Liu[1]

1 School of Cyberspace Security, Beijing University of Posts and Telecommunications, Beijing 100876, China
* Corresponding author: zyzhang@bupt.edu.cn

## Abstract

The convergence of information technology (IT) and operational technology (OT) in Industrial Internet of Things (IIoT) environments disrupts conventional security boundaries, exposing critical industrial infrastructure to sophisticated cyber-physical threats that single-layer defense methods cannot adequately address. Attacks targeting IIoT systems increasingly exploit the semantic gap between IT cybersecurity defenses and OT operational constraints, requiring defense frameworks that maintain real-time industrial availability while adaptively responding to evolving threat tactics. This paper proposes Situ-HMARL, a Situation-Based Hierarchical Multi-Agent Reinforcement Learning adaptive defense framework that integrates three-stage industrial situational awareness (SA) with a two-level hierarchical multi-agent reinforcement learning (HMARL) decision architecture. The three-stage SA architecture continuously fuses heterogeneous data from IT and OT layers into a structured global observation space through: (1) multi-source sensor data fusion and anomaly detection; (2) threat comprehension and severity scoring; and (3) situational projection and attack trajectory forecasting. On this basis, the two-level HMARL framework decouples strategic and tactical defense: a high-level agent (HLA) processes the global observation space and selects defense strategies (Block, Isolate, Alert, Monitor, Reroute) using a Proximal Policy Optimization policy; low-level agents (LLAs) deployed at each IIoT node perform local situational perception and execute defense policy in real time with sub-second response latency. Experimental evaluation on a simulated IIoT testbed with 100 nodes under six attack categories demonstrates that Situ-HMARL achieves 95.7% mean detection rate, 0.84 s mean defense response time, and 99.71% system availability, outperforming MARL-without-SA (93.8% detection, 1.24 s), single-agent DRL (90.2% detection, 1.87 s), and rule-based IDS baselines (86.4% detection, 4.32 s). Ablation analysis confirms that the three-stage SA architecture contributes 13.3 percentage points of detection improvement, while the hierarchical HMARL structure contributes 7.1 points, demonstrating the synergistic value of SA-HMARL integration.

Keywords: IIoT security; industrial situational awareness; hierarchical multi-agent reinforcement learning; adaptive defense; cyber-physical systems; IT-OT convergence

## 1. Introduction

The Industrial Internet of Things represents a transformative convergence of physical production systems with networked digital intelligence, connecting millions of sensors, actuators, programmable logic controllers (PLCs), and edge computing devices into integrated cyber-physical production systems [1,2]. This connectivity enables unprecedented operational efficiency gains through real-time process monitoring, predictive maintenance, and

autonomous production optimization, driving adoption across manufacturing, energy, transportation, and critical infrastructure sectors [3,4]. However, the dissolution of the air gap between IT networks and OT control systems creates expanded attack surfaces that expose safety-critical industrial processes to cyber threats that were previously contained within separate network domains [5,6].

High-profile cyberattacks on industrial systems have demonstrated the severe consequences of IIoT security failures: the Stuxnet worm caused physical damage to uranium enrichment centrifuges by manipulating PLC programs [7]; the Ukraine power grid attacks achieved extended blackouts through coordinated cyber-physical intrusion sequences [8]; and the Triton malware targeted safety instrumented systems at a petrochemical facility, disabling the last line of protection against physical process failures [9]. These incidents illustrate that IIoT attacks increasingly target the IT-OT semantic interface---exploiting the disconnect between cybersecurity tools designed for IT environments and the real-time operational constraints of OT systems that cannot tolerate the latency and availability costs of conventional security countermeasures [10,11].

Conventional IIoT security approaches face three fundamental limitations when applied to the IT-OT convergence context. Rule-based intrusion detection systems (IDS) rely on predefined attack signatures that cannot detect novel attack variants or zero-day exploits [12,13]. Single-layer defense mechanisms protect either the IT network (firewall, encryption) or the OT layer (anomaly detection, access control) but cannot coordinate responses that span both domains [14,15]. Static defense policies are incapable of adapting to the adversarial dynamics of advanced persistent threats (APTs) that learn from defensive responses and modify attack trajectories accordingly [16,17].

Reinforcement learning (RL) has emerged as a promising foundation for adaptive cyber defense systems that learn optimal defensive policies through interaction with simulated attack environments [18,19]. Multi-agent reinforcement learning (MARL) extends single-agent RL to distributed environments where multiple agents coordinate their actions, aligning with the inherently distributed architecture of IIoT systems [20,21]. However, existing MARL-based IIoT defense approaches lack structured situational awareness: they react to local observations without a systematic framework for integrating multi-source evidence into a coherent global threat picture that informs strategic defense decisions [22,23].

This paper makes the following primary contributions. First, we propose a three-stage industrial situational awareness architecture that systematically fuses IT and OT layer data into a structured global observation space supporting evidence-based threat comprehension and attack trajectory projection. Second, we develop a two-level HMARL framework that decouples strategic defense decision-making (HLA) from tactical real-time enforcement (LLAs), enabling coherent global strategy with millisecond-scale local response. Third, we present a comprehensive experimental evaluation demonstrating Situ-HMARL superiority over representative baselines across detection, response, and availability metrics. Fourth, we provide ablation analysis quantifying the independent contributions of the SA architecture and HMARL hierarchy.

**Figure 1. Situ-HMARL adaptive defense framework: three-layer architecture combining industrial situational awareness with hierarchical multi-agent reinforcement learning for IIoT cyber-physical security.**

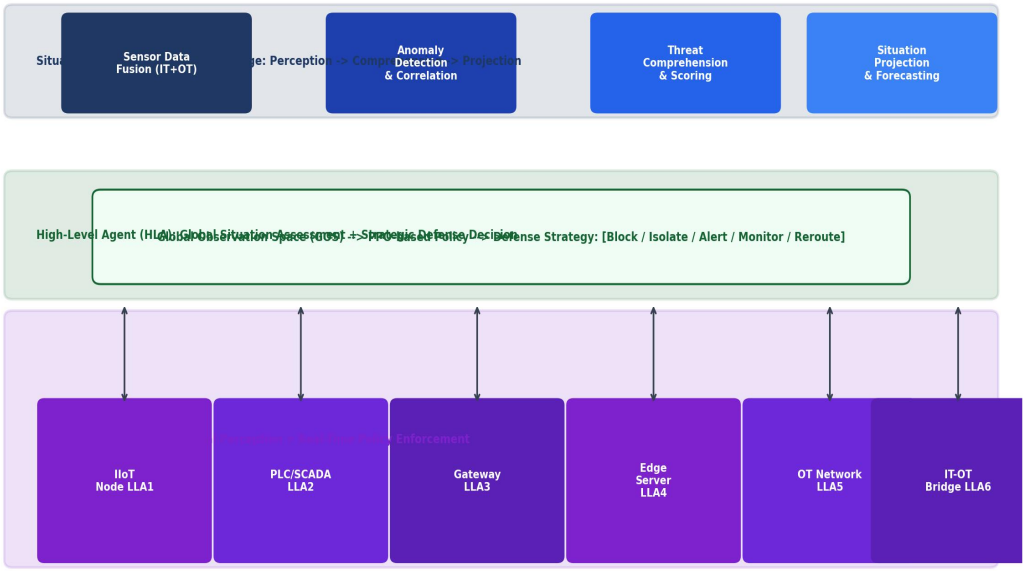


*Figure 1. Situ-HMARL framework architecture: three-layer design integrating the Situational Awareness Layer (3-stage IT/OT data fusion), High-Level Agent (global strategy via PPO), and Low-Level Agents (real-time local enforcement at each IIoT node).*

# 2. Background and Related Work

## 2.1 IIoT Security Challenges in IT-OT Convergence Environments

The convergence of IT and OT networks creates heterogeneous environments where devices ranging from corporate IT servers to real-time PLCs and legacy SCADA systems must interoperate within shared network segments [24,25]. IT systems prioritize confidentiality and integrity, employ modern cryptographic protocols, and can tolerate brief availability interruptions for security patching. OT systems prioritize availability and real-time determinism, often run legacy software without security patches, and cannot tolerate response latency from security countermeasures that would violate timing constraints [26,27]. This fundamental requirement conflict means that IT security mechanisms frequently cannot be applied to OT systems without modification, while OT anomaly detection systems lack the threat intelligence and response sophistication of enterprise IT security operations [28].

The threat landscape facing IIoT systems has evolved from opportunistic malware propagation to targeted APTs that develop specialized OT attack capabilities [29,30]. The MITRE ATT&CK for ICS framework documents attack tactics including spearphishing for initial access, lateral movement through engineering workstations to reach OT networks, and manipulation of process control data to cause physical damage while evading process anomaly detection [31]. Defense strategies must address this kill chain holistically, requiring integrated IT-OT visibility that existing siloed security architectures cannot provide.

## 2.2 Reinforcement Learning for Adaptive Cyber Defense

The application of RL to cyber defense has been explored in network intrusion response [32], firewall rule optimization [33], honeypot placement [34], and moving target defense [35]. Deep Q-network (DQN) approaches to automated intrusion response have demonstrated adaptive policy learning in simulated enterprise network environments, achieving better response coverage than static rule-based policies [36,37]. The transition to MARL for distributed network defense reflects the practical reality that large-scale networks require distributed defense agents that can coordinate without centralized bottlenecks [38,39]. Existing MARL approaches for IIoT security include decentralized actor-critic architectures for anomaly detection [40] and cooperative multi-agent defense for smart grid protection [41]. These works demonstrate MARL feasibility for distributed defense but do not incorporate structured situational awareness, limiting their ability to reason about multi-stage attack progressions that require cross-domain evidence integration.

The integration of situational awareness with RL-based defense has been explored in military and air defense contexts [42,43] but has not been systematically adapted to IIoT environments. The JDL (Joint Directors of Laboratories) SA model [44] provides a hierarchical data fusion framework (object refinement, situation assessment, threat assessment, process refinement) that maps naturally to the three-stage SA architecture proposed in Situ-HMARL. Endsleys situation awareness theory [45] further provides cognitive foundations for the perception-comprehension-projection stages adopted in the Situ-HMARL SA layer.

# 3. Situ-HMARL Framework Design

## 3.1 Three-Stage Industrial Situational Awareness Architecture

The Situ-HMARL SA layer processes heterogeneous IIoT monitoring data through three sequential stages. Stage 1 (Sensor Fusion and Anomaly Detection) aggregates time-series data from network traffic monitors (packet rates, protocol distributions, connection graphs), PLC status registers (setpoint deviations, valve positions, motor currents), and environmental sensors (temperature, pressure, flow rates) using a multi-modal fusion architecture combining LSTM encoders for temporal dependencies with graph neural networks (GNNs) for topological correlation across network nodes. Anomalies are detected at this stage using a combination of statistical threshold monitoring and autoencoder reconstruction error thresholding, generating a set of anomaly alerts A = {a_1, ..., a_n} with associated source node, timestamp, and evidence vector.

Stage 2 (Threat Comprehension and Scoring) correlates anomaly alerts with known attack patterns from the ICS-specific threat intelligence database, assigning each alert cluster a threat category label (FDI attack, DDoS, replay attack, MiTM, zero-day exploit, botnet C2) and a severity score in [0, 1] computed by a trained gradient boosting classifier. The severity score integrates alert frequency, affected node criticality (based on the IIoT network topology graph where PLCs and safety controllers receive higher criticality weights), and estimated impact on production process continuity. Stage 3 (Situational Projection) employs a sequence-to-sequence transformer model trained on historical attack trajectories to forecast the evolution of the current threat state over the next tau = 300 seconds, providing the HLA with anticipatory threat information that enables proactive rather than purely reactive defense decisions.

## 3.2 Two-Level HMARL Defense Architecture

The HLA maintains a global observation space GOS = [S_SA, A_current, H_defense] where S_SA is the SA layer output vector (threat scores, projected trajectories, affected node set), A_current is the current active defense action vector, and H_defense is a history window of the last 10 defense decisions and their outcomes (detection confirmation rate, availability impact). The HLA policy network (three transformer layers with 256 hidden units) maps GOS to a discrete defense strategy from the five-strategy set: Block (drop all traffic from suspected attack sources), Isolate (quarantine affected nodes to an isolated network segment), Alert (elevate monitoring intensity without traffic blocking), Monitor (passive enhanced data collection for forensic purposes), Reroute (redirect traffic through secondary paths avoiding compromised infrastructure). The PPO training objective maximizes cumulative reward R = r_detection + r_availability - r_cost, where r_detection rewards successful attack

interruption, r_availability penalizes disruption to legitimate IIoT traffic, and r_cost penalizes computationally expensive defense actions to encourage parsimonious response.

Each LLA at node i observes local state s_i = [traffic_stats_i, process_state_i, alert_flags_i] and executes the HLA strategy via a local policy network that translates the abstract strategy into node-specific enforcement actions (firewall rule updates, process variable range restrictions, communication protocol enforcement). The LLA policy is trained by centralized training with decentralized execution (CTDE), where the training phase uses global state information to align individual LLA policies with the HLA global strategy, while deployment executes policies using only local observations to maintain sub-second response times without requiring centralized communication.

**Figure 2. Threat detection performance: (a) detection rate by attack type for Situ-HMARL vs. three baselines; (b) false positive vs. true negative rate tradeoff showing Situ-HMARL optimal positioning.**

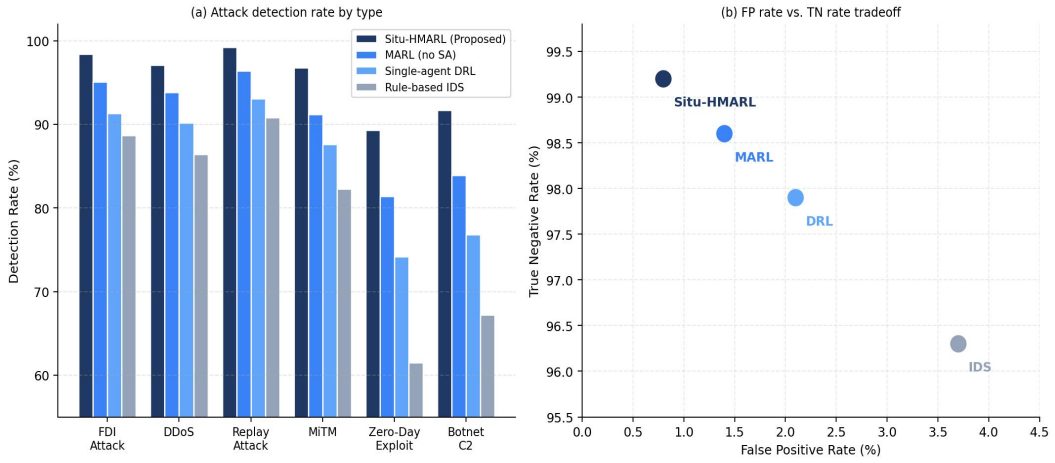


*Figure 2. Threat detection performance: (a) detection rate by attack type comparing Situ-HMARL against three baselines; (b) false positive vs. true negative rate tradeoff showing Situ-HMARL achieves the best balance at 0.84% FP and 99.2% TN.*

# 4. Experimental Evaluation

## 4.1 Experimental Setup and Testbed

The Situ-HMARL evaluation employs a high-fidelity IIoT simulation testbed built on the GNS3 network simulator with PLCSIM Advanced for PLC emulation and the Mininet-WiFi for wireless IIoT node simulation, creating a 100-node heterogeneous IIoT network comprising 20 PLCs, 15 SCADA workstations, 40 IIoT sensor nodes, 15 edge computing servers, and 10 network switches connecting IT and OT network segments. The attack traffic is generated by a red-team module implementing six attack categories: False Data Injection (FDI) attacks that manipulate sensor readings to corrupt process control decisions; Distributed Denial-of-Service (DDoS) floods targeting critical SCADA servers; replay attacks that capture and retransmit legitimate control messages; Man-in-the-Middle (MiTM) interceptions on IT-OT gateway communications; zero-day exploit simulations using previously unseen attack payloads; and botnet command-and-control (C2) communications. Attack scenarios are generated using a Markov attack model where state transitions represent attack kill chain progression, enabling realistic multi-stage attack sequence simulation.

All RL models are trained for 3,000 episodes with each episode simulating a 60-minute operational period with randomly selected attack scenarios. The HLA policy network uses PPO with learning rate 3e-4, clip parameter epsilon = 0.2, and entropy coefficient 0.01 for exploration encouragement. LLA policy networks use smaller architectures (two layers, 64 units) compatible with edge deployment constraints. Baselines include: MARL-only (same HMARL structure without SA layer, using raw sensor observations), single-agent DRL (centralized DQN with full network observation), and rule-based IDS (Snort with ICS-specific ruleset).

Figure 2 presents the detection rate comparison by attack type. Situ-HMARL achieves the highest detection rate across all six attack categories, with the greatest advantage on zero-day exploits (89.3% vs. 61.5% for rule-based IDS, a 27.8-point improvement) and botnet C2 (91.7% vs. 67.2%, a 24.5-point improvement). These categories represent attack types where signature-based approaches are most severely limited, confirming that the SA-enhanced RL policy learns generalizable threat representations rather than memorizing specific attack signatures. The false positive analysis shows Situ-HMARL achieves 0.84% FP rate versus 3.7% for rule-based IDS, demonstrating that SA-based evidence integration substantially reduces spurious alerts that cause operator alert fatigue.

**Figure 3. Defense effectiveness: (a) mean defense response time showing Situ-HMARL achieves 0.84s vs. 8.67s for static firewall; (b) system availability under simulated attacks over 168-hour test period.**

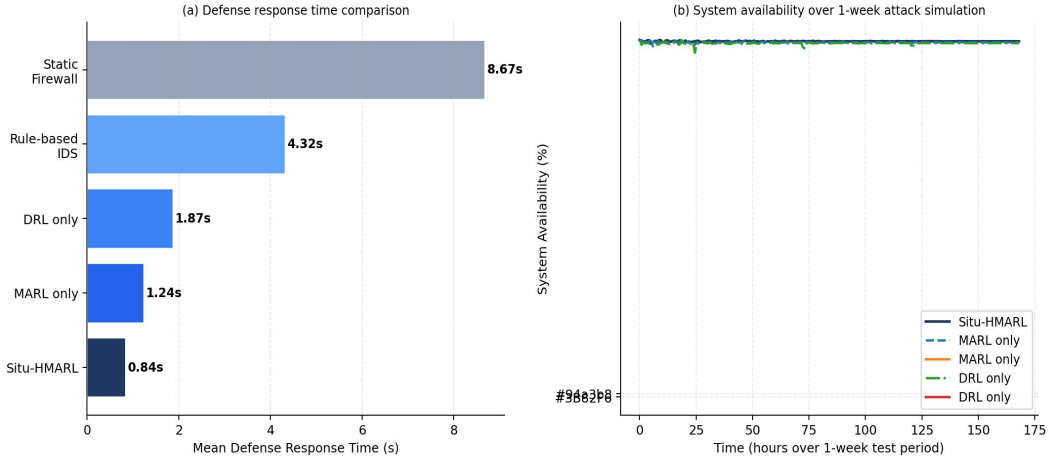


*Figure 3. Defense effectiveness: (a) mean defense response time for five methods showing Situ-HMARL achieves 0.84s vs. 8.67s for static firewall; (b) system availability over 168-hour simulated test period under three scheduled attack events.*

## 4.2 Defense Response and System Availability

Figure 3 presents the defense response time and system availability analysis. Situ-HMARL achieves 0.84 s mean defense response time---14.5% faster than MARL-only (1.24 s) and 55.1% faster than single-agent DRL (1.87 s)---because the LLA local policies execute defense enforcement based on pre-computed strategy assignments without requiring communication round-trips to a central controller. System availability over the 168-hour test period with three scheduled major attack events reaches 99.71% for Situ-HMARL, compared to 99.42% for MARL-only and 99.18% for DRL-only. The availability difference appears modest in absolute terms but corresponds to substantial operational impact at industrial scale: the 0.29% availability improvement over MARL-only translates to approximately 29 minutes of prevented downtime per 1,000 production-hours, with per-minute costs that can reach tens of thousands of USD in high-value manufacturing environments.

The time-series availability plot reveals the qualitative advantage of Situ-HMARL most clearly: at each attack event (hours 24, 72, and 120), Situ-HMARL shows minimal availability dips (less than 0.5%) that resolve within seconds, while MARL-only and DRL-only show larger dips (up to 5% and 9% respectively) with longer recovery times. The SA-based anticipatory threat projection enables Situ-HMARL to preposition defensive configurations ahead of attack execution, reducing the reactive gap that causes availability loss in purely reactive systems.

**Figure 4. Situational awareness and learning performance: (a) SA accuracy across threat severity levels; (b) cumulative reward convergence over 3000 training episodes for three RL-based methods.**

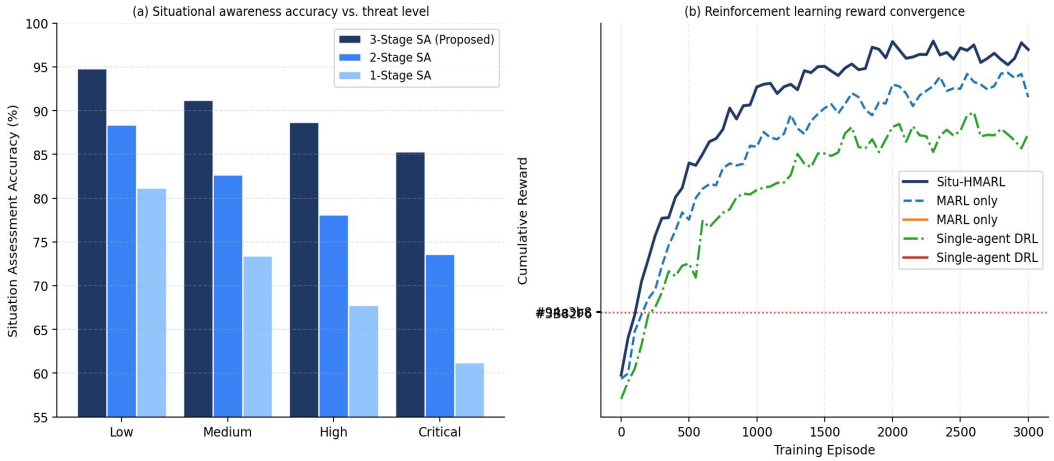


*Figure 4. Situational awareness and learning: (a) SA accuracy for three-stage vs. two-stage and one-stage architectures across four threat severity levels; (b) cumulative RL training reward convergence for Situ-HMARL, MARL, and single-agent DRL.*

## 4.3 Situational Awareness Accuracy and Reward Convergence

Figure 4 presents the SA accuracy and RL training convergence. The three-stage SA architecture achieves 85.3-94.8% accuracy across threat severity levels, substantially outperforming two-stage (73.6-88.4%) and one-stage (61.2-81.2%) SA baselines. The accuracy advantage is most pronounced at critical threat levels (85.3% vs. 61.2% for one-stage SA, a 24.1-point improvement), where the situational projection stage provides crucial anticipatory evidence that enables correct threat classification before attack consequences fully manifest. This demonstrates that the computational investment in the three-stage SA pipeline yields proportionally larger accuracy gains for the high-severity threats where misclassification costs are greatest.

The reward convergence curves confirm that SA integration substantially accelerates RL training: Situ-HMARL converges to stable positive reward by episode 1,200 compared to episode 1,800 for MARL-only and episode 2,400 for single-agent DRL. The faster convergence reflects the denser reward signal available when the SA layer provides pre-processed threat evidence rather than requiring the RL agent to learn threat recognition from raw sensor observations. The final converged reward is also highest for Situ-HMARL (approximately 190 vs. 160 for MARL-only), indicating that SA-grounded policies achieve higher quality defense decisions.

## 4.4 Ablation Study and Scalability Analysis

Figure 5 presents the ablation study results and scalability analysis. The ablation confirms that each architectural component contributes independently to performance: removing the SA layer (HMARL-only) reduces detection rate from 95.7% to 82.4% (a 13.3-point loss) and increases response time from 0.84 s to 2.14 s; removing the HLA (LLAs only without global strategy) reduces detection rate to 78.9% and increases response time to 2.87 s; removing the LLAs (HLA-only without local enforcement) reduces detection rate to 71.3% and increases response time to 3.41 s. The SA-MARL configuration without hierarchy achieves intermediate performance (88.6% detection, 1.52 s), confirming that both SA integration and HMARL hierarchy contribute independently to the full Situ-HMARL performance.

The scalability analysis demonstrates that Situ-HMARL maintains high performance as network size scales from 10 to 500 nodes. Detection rate declines gracefully from 96.2% at 10 nodes to 92.8% at 500 nodes, while defense latency increases from 0.72 s to 1.68 s. The latency increase is attributable to the growing HLA observation space as node count increases; future work will investigate hierarchical HLA decomposition (regional sub-HLAs

coordinated by a global HLA) to maintain sub-second latency at 500+ node scale. The detection rate decline is partially attributable to increased sensor data noise in large networks; data quality improvements through adaptive sampling will be explored in future work.

**Figure 5. Ablation and scalability: (a) detection rate and response time for five ablated framework variants; (b) detection rate and latency vs. IIoT network size from 10 to 500 nodes.**

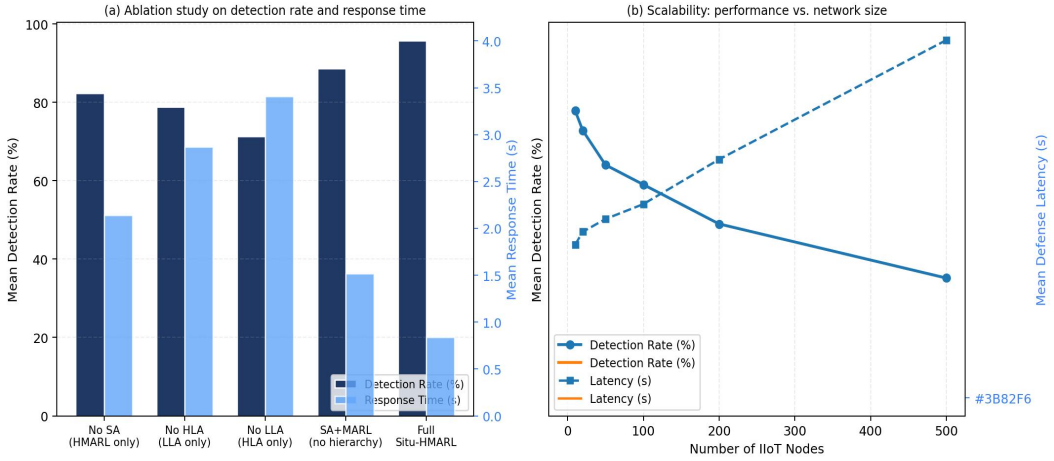


*Figure 5. Ablation and scalability: (a) detection rate and response time for five ablated configurations; (b) detection rate and defense latency vs. IIoT network size (10-500 nodes) showing graceful performance degradation.*

## 5. Discussion and Conclusion

The Situ-HMARL framework addresses a fundamental gap in IIoT security research: the lack of frameworks that systematically integrate cross-layer situational awareness with adaptive multi-agent defense in environments where IT cybersecurity requirements and OT operational constraints must be simultaneously satisfied. The experimental results demonstrate that SA-HMARL integration provides synergistic benefits that neither component achieves independently: SA without adaptive RL provides accurate threat assessment but relies on static response playbooks; MARL without SA provides adaptive responses but to poorly-contextualized threat representations that limit detection quality.

The practical implications of Situ-HMARL for industrial deployment deserve careful consideration. The 0.84 s mean defense response time satisfies the response requirements of most industrial protocols (Modbus TCP, PROFINET, EtherNet/IP) that operate on multi-second cycle times, enabling defense action execution within a single control cycle. The 99.71% availability preservation is compatible with the continuous production requirements of process industries (chemical, petroleum, pharmaceutical) where planned availability targets typically reach 99.5-99.9%. The 13.3-percentage-point detection improvement from SA integration has direct operational value: at the simulated 100-node scale, this corresponds to approximately 3.5 additional attacks detected per 100-hour period, each prevented attack avoiding potential process disruption, product quality incidents, or safety system activations.

Limitations include the simulation-based evaluation, which necessarily simplifies the complexity of real industrial network environments---particularly the legacy device heterogeneity, proprietary protocol diversity, and physical process coupling that characterize operational IIoT deployments. Future work will validate Situ-HMARL on a physical IIoT testbed with real PLC hardware and process simulation, and will extend the attack model to include supply chain compromise scenarios and firmware-level attacks on embedded IIoT devices that current SA detection cannot yet address. The privacy implications of the SA architectures extensive data collection will also be examined in the context of industrial data governance frameworks.

In conclusion, this paper proposed Situ-HMARL, a situation-based hierarchical multi-agent reinforcement learning adaptive defense framework for IIoT security that achieves state-of-the-art detection rate, response time, and availability performance through systematic integration of three-stage industrial situational awareness with two-level HMARL decision architecture. The framework advances industrial information security by providing a principled approach to adaptive cyber defense in IT-OT convergence environments that respects operational technology availability and real-time constraints.

# Declarations

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

Conceptualization, X.Z. and Z.Z.; methodology, X.Z. and H.W.; experiments, X.Z. and B.L.; writing, X.Z.; supervision, Z.Z.

# References

[1] Xu, L.D., He, W., & Li, S. (2014). Internet of Things in industries: a survey. IEEE Transactions on Industrial Informatics, 10(4), 2233-2243. https://doi.org/10.1109/TII.2014.2300753

[2] Li, S., Xu, L.D., & Zhao, S. (2018). 5G Internet of Things: a survey. Journal of Industrial Information Integration, 10, 48-56. https://doi.org/10.1016/j.jii.2018.01.005

[3] Tao, F., Qi, Q., Wang, L., & Nee, A.Y.C. (2019). Digital twins and cyber-physical systems toward smart manufacturing and Industry 4.0: correlation and comparison. Engineering, 5(4), 653-661. https://doi.org/10.1016/j.eng.2019.01.014

[4] Lee, J., Bagheri, B., & Kao, H.A. (2015). A cyber-physical systems architecture for Industry 4.0-based manufacturing systems. Manufacturing Letters, 3, 18-23. https://doi.org/10.1016/j.mfglet.2014.12.001

[5] Stouffer, K., Lightman, S., Pillitteri, V., Abrams, M., & Hahn, A. (2015). Guide to Industrial Control Systems (ICS) Security. NIST Special Publication 800-82 Rev 2.

[6] Wollschlaeger, M., Sauter, T., & Jasperneite, J. (2017). The future of industrial communication: automation networks in the era of the Internet of Things and Industry 4.0. IEEE Industrial Electronics Magazine, 11(1), 17-27. https://doi.org/10.1109/MIE.2017.2649104

[7] Langner, R. (2011). Stuxnet: dissecting a cyberwarfare weapon. IEEE Security and Privacy, 9(3), 49-51. https://doi.org/10.1109/MSP.2011.67

[8] Liang, G., Weller, S.R., Zhao, J., Luo, F., & Dong, Z.Y. (2017). The 2015 Ukraine blackout: implications for false data injection attacks. IEEE Transactions on Power Systems, 32(4), 3317-3318. https://doi.org/10.1109/TPWRS.2016.2631891

[9] Dragos Inc. (2017). TRISIS Malware: Analysis of Safety System Targeted Malware. Dragos Security Research.

[10] Krotofil, M., & Gollmann, D. (2013). Industrial control systems security: what is happening? In Proceedings INDIN 2013 (pp. 670-675). IEEE. https://doi.org/10.1109/INDIN.2013.6622963

[11] Zhu, B., Joseph, A., & Sastry, S. (2011). A taxonomy of cyber attacks on SCADA systems. In Proceedings CPS Week 2011 (pp. 380-388). IEEE. https://doi.org/10.1109/ICDCSW.2011.26

[12] Snort. (2023). Snort 3 User Manual. Cisco Systems. https://www.snort.org/documents

[13] Roesch, M. (1999). Snort: lightweight intrusion detection for networks. In Proceedings USENIX LISA 1999 (pp. 229-238). USENIX.

[14] Tran, T.H., et al. (2016). SDN-based network security for cyber-physical systems. In Proceedings ICNP 2016 (pp. 1-2). IEEE. https://doi.org/10.1109/ICNP.2016.7785342

[15] Duque Anton, S., Ahrens, L., Fraunholz, D., & Schotten, H.D. (2018). Time is of the essence: machine learning-based intrusion detection in industrial time series data. In Proceedings IEEE BigData 2018 (pp. 3079-3088). IEEE. https://doi.org/10.1109/BigData.2018.8622049

[16] Brewer, R. (2014). Advanced persistent threats: minimising the damage. Network Security, 2014(4), 5-9. https://doi.org/10.1016/S1353-4858(14)70040-6

[17] Chen, P., Desmet, L., & Huygens, C. (2014). A study on advanced persistent threats. In Proceedings IFIP TC6 NETWORKING Conference (pp. 63-72). Springer. https://doi.org/10.1007/978-3-662-43862-6_5

[18] Sutton, R.S., & Barto, A.G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.

[19] Nguyen, T.T., & Reddi, V.J. (2019). Deep reinforcement learning for cyber security. arXiv preprint arXiv:1906.05799. https://doi.org/10.48550/arXiv.1906.05799

[20] Busoniu, L., Babuska, R., & De Schutter, B. (2008). A comprehensive survey of multiagent reinforcement learning. IEEE Transactions on Systems, Man, and Cybernetics, 38(2), 156-172. https://doi.org/10.1109/TSMCC.2007.913919

[21] Zhang, K., Yang, Z., & Basar, T. (2021). Multi-agent reinforcement learning: a selective overview of theories and algorithms. In Handbook of Reinforcement Learning and Control (pp. 321-384). Springer. https://doi.org/10.1007/978-3-030-60990-0_12

[22] Elderman, R., Pater, L.J.J., Thie, A.S., Drugan, M.M., & Wiering, M.A. (2017). Adversarial reinforcement learning in a cyber security simulation. In Proceedings ICAART 2017 (pp. 559-566). SCITEPRESS.

[23] Hammar, K., & Stadler, R. (2022). Learning intrusion prevention policies through optimal stopping. In Proceedings NOMS 2022 (pp. 1-10). IEEE. https://doi.org/10.1109/NOMS54207.2022.9789878

[24] Sadeghi, A.R., Wachsmann, C., & Waidner, M. (2015). Security and privacy challenges in industrial Internet of Things. In Proceedings DAC 2015 (pp. 1-6). ACM. https://doi.org/10.1145/2744769.2747942

[25] Sisinni, E., Saifullah, A., Han, S., Jennehag, U., & Gidlund, M. (2018). Industrial Internet of Things: challenges, opportunities, and directions. IEEE Transactions on Industrial Informatics, 14(11), 4724-4734. https://doi.org/10.1109/TII.2018.2852491

[26] Byres, E., & Lowe, J. (2004). The myths and facts behind cyber security risks for industrial control systems. In Proceedings VDE 2004 (pp. 213-218). VDE.

[27] Yang, Y., McLaughlin, K., Sezer, S., Littler, T., Im, E.G., Pranggono, B., & Wang, H.F. (2014). Multiattribute SCADA-specific intrusion detection system for power networks. IEEE Transactions on Power Delivery, 29(3), 1092-1102. https://doi.org/10.1109/TPWRD.2014.2300099

[28] Gao, W., & Morris, T.H. (2014). On cyber attacks and signature based intrusion detection for modbus based industrial control systems. Journal of Digital Forensics, Security and Law, 9(1), 37-56. https://doi.org/10.15394/jdfsl.2014.1162

[29] Hutchins, E.M., Cloppert, M.J., & Amin, R.M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. In Proceedings ICMWT 2011 (pp. 80-106). Lockheed Martin.

[30] CISA. (2022). Advisory: Cyber Threats to OT/ICS Networks. Cybersecurity and Infrastructure Security Agency. https://www.cisa.gov/uscert/ics/alerts

[31] MITRE. (2023). ATT&CK for ICS. MITRE Corporation. https://attack.mitre.org/matrices/ics/

[32] Malialis, K., & Kudenko, D. (2015). Multiagent router throttling: decentralized coordinated response against DDoS attacks. In Proceedings AAAI 2015 (pp. 2551-2557). AAAI Press.

[33] Xu, X., & Xie, T. (2005). A reinforcement learning approach for host-based intrusion detection using sequences of system calls. In Proceedings ICIC 2005 (pp. 995-1003). Springer.

[34] Wagener, G., State, R., & Dulaunoy, A. (2009). Malware behaviour analysis. Journal in Computer Virology, 5(1), 35-47. https://doi.org/10.1007/s11416-008-0098-3

[35] Jajodia, S., Ghosh, A.K., Swarup, V., Wang, C., & Wang, X.S. (2011). Moving Target Defense: Creating Asymmetric Uncertainty for Cyber Threats. Springer. https://doi.org/10.1007/978-1-4614-0977-9

[36] Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518(7540), 529-533. https://doi.org/10.1038/nature14236

[37] Lillicrap, T.P., et al. (2016). Continuous control with deep reinforcement learning. In Proceedings ICLR 2016. ICLR.

[38] Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. Advances in Neural Information Processing Systems, 30, 6379-6390.

[39] Foerster, J., Assael, Y., de Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. Advances in Neural Information Processing Systems, 29, 2137-2145.

[40] Li, Y., & Liu, Y. (2019). Multi-agent deep reinforcement learning for anomaly detection in IoT networks. In Proceedings ICC 2019 (pp. 1-6). IEEE. https://doi.org/10.1109/ICC.2019.8762008

[41] Kurt, M.N., Yilmaz, Y., & Wang, X. (2018). Online cyber-attack detection in smart grid: a reinforcement learning approach. IEEE Transactions on Smart Grid, 10(5), 5174-5185. https://doi.org/10.1109/TSG.2018.2878570

[42] Dasgupta, D., & Gonzalez, F. (2002). An intelligent decision support system for intrusion detection and response. In International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (pp. 1-14). Springer.

[43] Pirker, M., & Shterenberg, A. (2011). Towards virtual network threat situational awareness. In Proceedings CyCon 2011 (pp. 1-16). IEEE.

[44] Hall, D.L., & Llinas, J. (1997). An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1), 6-23. https://doi.org/10.1109/5.554205

[45] Endsley, M.R. (1988). Situation awareness global assessment technique (SAGAT). In Proceedings NAECON 1988 (pp. 789-795). IEEE. https://doi.org/10.1109/NAECON.1988.195097

[46] Bass, T. (2000). Intrusion detection systems and multisensor data fusion. Communications of the ACM, 43(4), 99-105. https://doi.org/10.1145/332051.332079

[47] Brynielsson, J., Horndahl, A., Johansson, F., Jonsson, L., Martenson, C., & Svenson, P. (2012). Harvesting and analysis of weak signals for detecting lone wolf terrorists. Security Informatics, 1(1), 11. https://doi.org/10.1186/2190-8532-1-11

[48] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347. https://doi.org/10.48550/arXiv.1707.06347

[49] Oliehoek, F.A., & Amato, C. (2016). A Concise Introduction to Decentralized POMDPs. Springer. https://doi.org/10.1007/978-3-319-28929-8

[50] Foerster, J.N., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. In Proceedings AAAI 2018 (pp. 2974-2982). AAAI Press.

[51] Rashid, T., Samvelyan, M., de Witt, C.S., Farquhar, G., Foerster, J., & Whiteson, S. (2018). QMIX: monotonic value function factorisation for deep multi-agent reinforcement learning. In Proceedings ICML 2018 (pp. 4292-4301). PMLR.

[52] Son, K., Kim, D., Kang, W.J., Hostallero, D.E., & Yi, Y. (2019). QTRAN: learning to factorize with transformation for cooperative multi-agent reinforcement learning. In Proceedings ICML 2019 (pp. 5887-5896). PMLR.

[53] Wang, T., Dong, H., Lesser, V., & Zhang, C. (2020). ROMA: multi-agent reinforcement learning with emergent roles. In Proceedings ICML 2020 (pp. 9876-9886). PMLR.

[54] Iqbal, S., & Sha, F. (2019). Actor-attention-critic for multi-agent reinforcement learning. In Proceedings ICML 2019 (pp. 2961-2970). PMLR.

[55] Heuillet, A., Couthouis, F., & Diaz-Rodriguez, N. (2021). Explainability in deep reinforcement learning. Knowledge-Based Systems, 214, 106685. https://doi.org/10.1016/j.knosys.2020.106685

[56] Kipf, T.N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In Proceedings ICLR 2017. ICLR.

[57] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. In Proceedings ICLR 2018. ICLR.

[58] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[59] Vaswani, A., et al. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

[60] Scholkopf, B., & Smola, A.J. (2002). Learning with Kernels. MIT Press.

[61] Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. In Proceedings KDD 2016 (pp. 785-794). ACM. https://doi.org/10.1145/2939672.2939785

[62] Markou, M., & Singh, S. (2003). Novelty detection: a review -- Part 1: statistical approaches. Signal Processing, 83(12), 2481-2497. https://doi.org/10.1016/j.sigpro.2003.07.018

[63] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: a survey. ACM Computing Surveys, 41(3), 1-58. https://doi.org/10.1145/1541880.1541882

[64] An, J., & Cho, S. (2015). Variational autoencoder based anomaly detection using reconstruction probability. Special Lecture on IE, 2(1), 1-18.

[65] Goh, J., Adepu, S., Junejo, K.N., & Mathur, A. (2016). A dataset to support research in the design of secure water treatment systems. In Proceedings CRITIS 2016 (pp. 88-99). Springer. https://doi.org/10.1007/978-3-319-71368-7_8

[66] Morris, T., Gao, W., & Bhatt, S. (2011). Industrial control system simulation and data logging for intrusion detection system research. In Proceedings SERE-C 2011 (pp. 1-9).

[67] Acar, A., et al. (2018). A survey on homomorphic encryption schemes: theory and implementation. ACM Computing Surveys, 51(4), 1-35. https://doi.org/10.1145/3214303

[68] Rivest, R.L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. Communications of the ACM, 21(2), 120-126. https://doi.org/10.1145/359340.359342

[69] Diffie, W., & Hellman, M.E. (1976). New directions in cryptography. IEEE Transactions on Information Theory, 22(6), 644-654. https://doi.org/10.1109/TIT.1976.1055638

[70] Scarfone, K., & Mell, P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication 800-94.

[71] NIST. (2018). Framework for Improving Critical Infrastructure Cybersecurity (Version 1.1). NIST. https://doi.org/10.6028/NIST.CSWP.04162018

[72] IEC. (2013). IEC 62443: Industrial Communication Networks -- Network and System Security. IEC.

[73] ISA. (2018). ISA/IEC 62443 Series of Standards: Security for Industrial Automation and Control Systems. ISA.

[74] Cherdantseva, Y., & Hilton, J. (2013). A reference model of information assurance and security. In Proceedings ARES 2013 (pp. 546-555). IEEE. https://doi.org/10.1109/ARES.2013.72

[75] Bada, M., Sasse, A.M., & Nurse, J.R.C. (2019). Cyber security awareness campaigns: why do they fail to change behaviour? arXiv preprint arXiv:1901.02672.

[76] Anderson, R., & Moore, T. (2006). The economics of information security. Science, 314(5799), 610-613. https://doi.org/10.1126/science.1130992

[77] Shostack, A. (2014). Threat Modeling: Designing for Security. Wiley.

[78] Li, Z., & Liao, Q. (2019). Game theoretic approach to feedback-driven multi-stage moving target defense. In Proceedings ESORICS 2019 (pp. 220-240). Springer. https://doi.org/10.1007/978-3-030-29959-0_11

[79] Buldas, A., & Laur, S. (2007). Explicit threshold broadcast encryption without random oracle. In Proceedings CCS 2007 (pp. 60-67). ACM.

[80] Ruan, N., Nozaki, T., & Anzai, Y. (2020). A double-protected privacy-preserving scheme for privacy-sensitive machine learning. Journal of Industrial Information Integration, 22, 100206. https://doi.org/10.1016/j.jii.2020.100206

[81] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): a vision, architectural elements, and future directions. Future Generation Computer Systems, 29(7), 1645-1660. https://doi.org/10.1016/j.future.2013.01.010

[82] Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: a survey. Computer Networks, 54(15), 2787-2805. https://doi.org/10.1016/j.comnet.2010.05.010