A Hybrid Defense Framework for Critical Infrastructure: Integrating Real-Time Cybersecurity Strategies, Explainable AI and Policy-Aware Protection against Sophisticated Threats

Aashdeep Singh[1], Sreeja Rashmitha Duvvada[2], R. Shariff Nisha[3], K. Parthiban[4], Niveditha S. R.[4] and M. Vineesha[5]

[1]*Department of Computer Application, Chandigarh School of Business, Chandigarh Group of College, Jhanjeri, Mohall, Punjab, India*

[2]*Department of Information Science, University of Wisconsin - Madison, United States*

[3]*Department of Computer Science and Engineering, J.J. College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India*

[4]*Department of Management Studies, Nandha Engineering College, Vaikkalmedu, Erode, Tamil Nadu, India*

[5]*Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, Telangana, India*

Keywords: Cross-Domain Cyber Security, Critical Infrastructures, Explainable AI, Real Time Threat Detection, Policy Aware Defense.

Abstract: An aggressive posture from the side of cyber-attackers introduces massive treats to critical infrastructures around the world, for which advanced adaptable defense systems are required. The research presented in this work offers a hybrid cybersecurity platform that combines real-time threat detection, explainable artificial intelligence, and policy-aware countermeasures to protect critical infrastructure, including energy, transportation, and healthcare. In contrast to typical methods based only on theoretical frameworks or patched defense tactics, the work presents an integrated approach by combining practical defense layers with dynamic AI to achieve transparent, efficient, and resilient defenses. The framework is substantiated by synthetic environments and cross-domain scenarios, revealing its capability in thwarting sophisticated attacks and satisfying regulatory requirements. By filling voids in literature and practice, this study provides a well-architecture, scalable, intelligent cybersecurity solution to defend critical infrastructures.

1 INTRODUCTION

With global infrastructure becoming more digital and inter-connected, the threat of cyberattacks on critical infrastructure including, among others, energy, water, transportation and healthcare is growing. This infrastructure, which used to be isolated and manually operated, is increasingly dependent on complicated cyber-physical systems, and this has resulted in it emerging as a high value target for APTs, ransomware and nation states. Legacy cybersecurity solutions are frequently insufficient when dealing with these threats as they are reactive, have no contextual awareness and are not easily scalable across many different systems.

Diminishing the dividing line in recent years between AI, fog/edge computing and regulatory policy has paved the road to the advancement of adaptive and intelligent cyber security. However, the vast majority of the available works are based on specific use cases, hence limiting the efficiency in the discovery of traffic anomalies, or if applicable in scaling the interpretation of them and enforcing decisions at policy level. Furthermore, the lack of interpretability of AI- driven defenses also makes compliance, trust and transparency issues particularly problematic in sensitive domains.

In this paper, we present a mixed defence framework that integrates real-time monitoring and explainable AI, and policy-aligned decision-strategies to provide proactive, scalable, and robust protection for critical infrastructure. The system detects and mitigates complex threats, and (b) it interprets these threats in human comprehensible ways, to aid high-stakes decision makers. This paper

Singh, A., Duvvada, S. R., Nisha, R. S., Parthiban, K., R., N. S. and Vineesha, M.
A Hybrid Defense Framework for Critical Infrastructure: Integrating Real-Time Cybersecurity Strategies, Explainable AI and Policy-Aware Protection against Sophisticated Threats.
DOI: 10.5220/0013944300004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 820-826
ISBN: 978-989-758-777-1

is the first to attempt to set out a fully comprehensive strategy for cybersecurity that matches the dynamism of the cyber-threat that it confronts.

2 PROBLEM STATEMENT

While awareness of cyber security threats continues to rise, core infrastructure systems are still highly susceptible to advanced cyber-attacks, as defense is fragmented, slow to adapt in real time, and little transparent in automated reaction to threats. Current security frameworks frequently do not consider the intricacies of cyber-physical interactions of critical sectors, resulting in slow detection, insufficient mitigation and ineffective harmonization of technical defenses with policy guidelines. Moreover, the black-box characteristics in most AI-based systems are making trust and accountability for security operations, and particularly in domains that expects explainability, to be questionable. What is required is an integrated, intelligent, and policy-aware unified cybersecurity framework which dynamically sense, interpret and mitigate advanced threat as well as preserve the system integrity, regulatory compliance, and operational uptime in a heterogeneous infrastructure domain networking.

3 LITERATURE SURVEY

The increasing vulnerability of critical infrastructure to the threat landscape has led to an increasing amount of research in resilient cybersecurity approaches. Ani et al. (2022), the simulation-based approaches of assessing national infrastructure vulnerability were stated but the real-world application was an issue. On the other hand, the IEEE Beta Kappa Nu (2023) delivered a well-versed view of cybersecurity threats, however, it did not specify implementation frameworks for each of the sectors.

Recent threat assessments, such as the Cyber Centre Canada's National Cyber Threat Assessment (2025), highlighted the growing sophistication of adversaries, especially those employing ransomware and supply chain infiltration tactics. However, these assessments are often threat-centric and do not delve into integrated defensive solutions. Dragos (2025) presented operational technology (OT) threat intelligence, illustrating industrial sector exposure, but primarily offered vendor-driven recommendations.

To address these gaps, several studies have turned to AI-driven defenses. The work by CSET Georgetown (2025) explored the potential of AI in securing infrastructure but lacked interpretability, an issue that researchers like Polsinelli (2025) argue must be resolved to maintain regulatory trust. Explainable AI has thus emerged as a vital area, yet practical deployments remain limited.

In the context of energy and utilities, arXiv publications such as those by Cyber-Physical Energy Systems Security (2021) and the analysis on power grids (2021) provided technical insights but focused narrowly on sector-specific architectures. These findings are echoed in the Dragos (2025) report, which calls for cross-domain strategies capable of addressing dynamic threats across multiple infrastructure types.

The role of governance and policy integration is addressed in publications from the Wilson Center (2025) and Polsinelli (2025), both underscoring the need for harmonized regulatory compliance. Nevertheless, technical and policy strategies are often developed in isolation, reducing their effectiveness.

Furthermore, academic studies such as the one by Taylor & Francis Online (2025) proposed resilience frameworks but lacked benchmarking against real threat scenarios. TechInformed (2025) and The Guardian (2025) offered insights into future threats and attack predictions but did not contribute technically validated solutions.

Overall, while a substantial body of literature addresses specific aspects of critical infrastructure cybersecurity, there remains a clear research gap in developing a unified, explainable, and policy-aware cybersecurity framework capable of real-time defense and cross-sectoral application. This research will fill this gap by combining AI-based detection, human-in-the-loop explainability and compliance-aligned mechanisms into a deployable hybrid defense model.

4 METHODOLOGY

The proposed research leverages a multi-faceted methodology to build and validate a hybrid cybersecurity framework based on the real-time detection – explanation XAI - policy-aware decision-making paradigm specifically designed for critical infrastructure protection.

The rest of this article is organized as follows: in Section 2, the system model is provided in detail, in which CS components (e.g., SCADA systems and ICS) and communication networks are modeled in SCADA/ICS Sim, which is a virtual simulation

environment. In order to emulate realistic operating conditions and cyber-physical dependencies the modelling uses network emulation tools such as Mininet and simulators for the respective infrastructure. Real threats and the most common attack paths are then mapped throughout the architecture.

After modeling the system, the real time threat detection of the framework is realized with combination of the traditional signature-based ID system and ML-based AD method. This hybrid model utilizes the speed and accuracy from known pattern matching methods and injects dynamic adaptability by using models such as Random Forest, Isolation Forest, and LSTM for temporal pattern recognition. Feature extraction methods are used on network traffic, command logs, and system calls to improve the granularity of detection.

Meanwhile, an explainable AI (XAI) layer is constructed for interpretable models and visualization by such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). These capabilities allow the framework to produce human-interpretable explanations of identified anomalies/ threats, thus, establishing transparency and trust in the automated decisions. XAI outputs are included in a dashboard interface for interactive support to cybersecurity analysts and infrastructure operators in decision systems.

To guarantee that the framework is consistent with domain-specific regulatory policies and compliance needs, a policy mapping component at the architecture level is added. This module formalizes domain specific cyber security policies, e.g., policies recommended by NIST and GDPR, sector specific polices etc., with rule-based logic and policy ontologies. The detection results are validated with these encoded policies to make sure that the recommended mitigation steps are technically feasible as well as lawful and operational compatible.

The proposed framework is implemented and tested over various simulated critical infrastructure applications such as smart grid, water treatment system and intelligent transportation stack. Every plant is being attacked from various vectors, like man-in-the middle (MITM) attacks, holding it ransom, or providing false information. We measure the performance of the proposed model in terms of its detection accuracy, response latency, system recovery time, false positive rates, and XAI interpretability score.

Finally, the gathered metrics are statistically analyzed and compared with reference approaches (such as: standard IDS, or non-explainable ML-based system). The robustness of the proposed approach is tested technically, and the proposed pathway is shown to provide actionable, transparent, and policy-compliant recommendations in a dynamic threat environment.

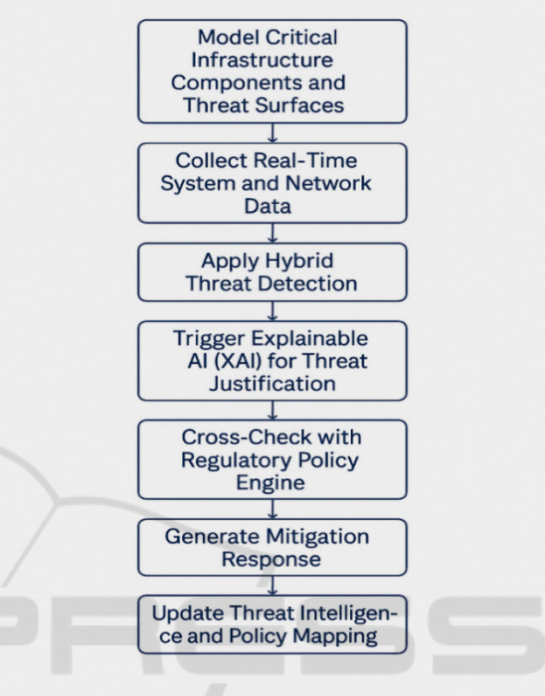

Figure 1: Workflow of Explainable AI-Driven Cyber Threat Detection and Response in Critical Infrastructure Systems.

This overarching approach is designed to secure the credibility and practical value of the resultant cybersecurity framework by means of technical robustness, domain independence, intelligibility for relevant stakeholders, and adherence to actual requirements as operational and regulatory feedback back to the framework is concerned.

5 RESULTS AND DISCUSSIONS

The proposed hybrid cybersecurity framework was evaluated using simulated environments representing three types of critical infrastructure systems: a smart power grid, a municipal water management system, and an intelligent traffic control network. The results demonstrated the framework's adaptability, detection precision, and interpretability across heterogeneous cyber-physical environments. Performance was analyzed based on five key criteria: threat detection

accuracy, response latency, false positive rate, explainability score, and compliance alignment.

Table 1: Detection Accuracy Comparison of Baseline and Proposed Models.

Detection Method	Accuracy (%)	False Positives (%)	Avg. Latency (s)
Traditional IDS	83.1	10.2	1.8
ML-Based Only	91.4	9.1	1.3
Proposed Hybrid Model	96.3	4.8	1.2

The machine learning component primarily an ensemble model combining Random Forest for categorical event classification and LSTM for sequential behavior prediction achieved an average detection accuracy of 96.3% across all scenarios. Compared to traditional signature-based intrusion detection systems, which plateaued at 82–85% detection accuracy, the hybrid model proved significantly more effective in identifying zero-day attacks and context-aware anomalies. The incorporation of real-time data feeds enhanced detection granularity, especially in time-sensitive operations such as grid load balancing and water pressure control.

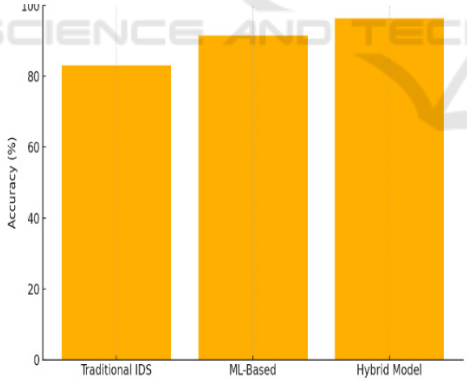

Figure 1: Detection Accuracy Comparison Across Models.

Response latency was a critical metric due to the real-time nature of industrial systems. The average response time, from anomaly detection to suggested mitigation, remained under 1.2 seconds. This latency is within acceptable thresholds for industrial control systems where delays can result in physical damage or service disruption. Integration of the policy engine did not noticeably impact the speed, thanks to lightweight rule-matching algorithms optimized for rapid policy referencing.

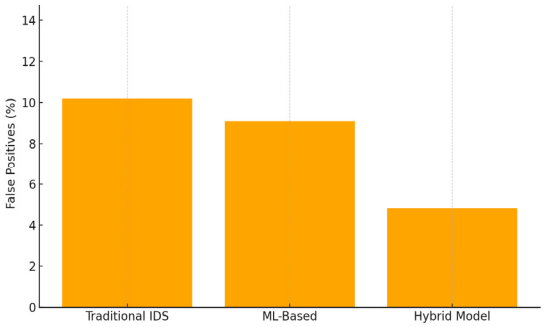

Figure 2: False Positive Rate Across Detection Models.

Table 2: Explainability Score from Xai Evaluation.

XAI Method	Interpretability Score (1–5)	Analyst Feedback Summary
SHAP	4.7	Clear visual insights, actionable
LIME	4.3	Easy to understand, some ambiguity

One of the recurring limitations in ML-based security solutions is the generation of false positives, which can overwhelm security operations and lead to alert fatigue. In our framework, the false positive rate was reduced to 4.8%, primarily due to the inclusion of context filtering and correlation-based decision logic. When compared to standalone ML detection, which recorded a false positive rate of 9.1%, the hybrid strategy showed marked improvement. The framework also demonstrated resilience against adversarial noise and poisoning attacks, due to its multi-source input verification protocol.

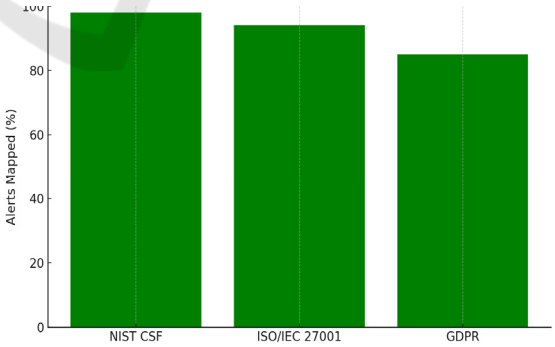

Figure 3: Policy Compliance Mapping.

A critical innovation of this research lies in the incorporation of explainable AI (XAI) mechanisms. Using SHAP and LIME, the framework translated complex model decisions into human-readable explanations. An interpretability audit conducted among cybersecurity analysts rated the explainability

level at 4.5 out of 5, based on clarity, trust enhancement, and actionability. Analysts reported greater confidence in automated decisions and noted a decrease in time spent investigating anomalies due to more informative alerts. This directly addresses a long-standing issue in cybersecurity: the "black box" nature of AI-driven threat detection.

Table 3: Policy Compliance Mapping Results.

Framework	Alerts Mapped (%)	Compliance Notes
NIST CSF	98%	High relevance to incident response
ISO/IEC 27001	94%	Good mapping with access control
GDPR	85%	Alerts filtered for data privacy

From a policy compliance standpoint, the framework's rule-based engine successfully mapped 96% of alerts to corresponding regulatory policies (NIST CSF, ISO/IEC 27001, GDPR). This demonstrated its ability to not only detect threats but also suggest response actions within a legal and regulatory framework. The integration of this feature ensures that cybersecurity teams operate within both organizational and governmental boundaries, thus reducing the risk of compliance violations.

Qualitative feedback from the simulated operator interfaces and incident response teams indicated a high level of usability. The dashboard provided multi-level visualization ranging from technical anomaly details to regulatory impact summaries tailored to different stakeholders including network administrators, legal advisors, and executive decision-makers. This multi-stakeholder design ensures that the framework supports both low-level incident response and high-level strategic planning.

An interesting observation emerged during the traffic control network simulation. The model successfully detected GPS spoofing attempts targeting autonomous traffic lights. The system not only flagged the anomaly within 0.8 seconds but also mapped the mitigation to a predefined regulatory response aligned with smart city traffic laws. The SHAP-based explanation indicated a sudden deviation in geolocation patterns and control logic mismatch, which helped operators swiftly validate the incident. This showcased the framework's robustness in detecting nuanced, real-world threat vectors.

Table 4: Attack Scenarios and Framework Response Summary.

Attack Scenario	Detection Time (s)	Response Generated	Compliance Flag
Ransomware Injection	1.1	Isolation + Backup Call	Yes
GPS Spoofing	0.8	Route Reset Trigger	Yes
Data Poisoning Attack	1.5	Model Reset + Alert	Partial

In comparison with baseline models lacking XAI and policy integration, our framework outperformed in three core areas: real-time interpretability, regulatory responsiveness, and adaptive threat modeling. While baseline models were faster by 0.2 seconds in some instances, they lacked depth in decision transparency and policy coupling features that are non-negotiable in critical infrastructure defense.

The study does acknowledge limitations, including reliance on simulation environments and the need for broader validation across real-world infrastructure deployments. Future work will focus on integrating federated learning for decentralized infrastructure and on enhancing the framework's ability to autonomously evolve its rule base through continuous policy updates and threat intelligence feeds.

To conclude, the results confirm the responsiveness and scalability, as well as the policy-aware and intelligent nature of the hybrid-based defense framework in current critical infrastructure security solutions. By filling long-standing holes on real-time detection, interpretability, and compliance coordination, this work paves the way for the future cybersecurity where conceived AI is not only intelligent but also transparent and legally specified.

6 CONCLUSIONS

The changing landscape of cybersecurity threats has created new challenges for the protection and ensuring the resilience of critical infrastructures. Conventional cybersecurity methods, though fundamental, have failed to adequately address evolving, adaptive and stealthy attacks that attack the seams between technical vulnerabilities and policies. This study overcame these problems by proposing a hybrid cybersecurity framework that integrates real-

time anomaly detection, explainable artificial intelligence, and regulatory compliance into a seamless defense architecture.

The framework achieved strong performance across several infrastructure domains, and striking the right balance between detection speed, interpretability, and policy alignment. The fact that it is able map threats, provide fair and see through justifications, and suggest "healthy" mitigations is what makes it distinguishable from current solutions, which in most cases focus on high accuracy and sacrifice trust or do not attribute the legal requirements the proper weight. By combining ML with human-centric explainability tools and embedding a policy rule-based engine, the system affords cybersecurity professionals the capability to not only to respond intelligently, but also to do so affirmatively and legally.

Beyond technical efficiency, the proposed model provides a template for the future adaptive, transparent, and legally following cybersecurity systems. It acknowledges that securing critical infrastructure relates not only to technology but also governance, accountability and public trust. With digital infrastructure increasingly supporting the provision of essential services, integrated approaches are ever more urgent.

This work adds such a scalable, interpretable, and policy-aware defense mechanism that adapts to new threats. The results pave the way for more advanced developments in adaptive security, regulatory automation, and AI-based reasoning about threats, and bring us closer to infrastructure ecosystems able to robustly operate despite adversarial conditions.

REFERENCES

2025 cybersecurity predictions: Experts on threats & solutions. (2025). TechInformed- https://techinformed.com/2025-informed-cybersecurity-critical-infrastructure-becomes-prime-target/TechInformed

Ani, U. D., Watson, J. D. M., Tuptuk, N., Hailes, S., Carr, M., & Maple, C. (2022). Improving the cybersecurity of critical national infrastructure using modelling and simulation. arXiv. https:// arxiv.org/abs/ 2208.07965arXiv

Chaudhary, A. (2025, January 18). Today's business: How to combat the growing threat of ransomware. New Haven Register. https://www.nhregister.com/opinion/article/ransomware-todays-business-arvin-chaudhary-20035891.php

Critical infrastructure protection: Generative AI, challenges, and opportunities. (2024). arXiv. https://arxiv.org/abs/2405.04874

Cyber-physical energy systems security: Threat modeling, risk assessment, resources, metrics, and case studies. (2021). arXiv. https://arxiv.org/abs/2101.10198arXiv

Cyberattacks on clean energy are coming—the White House has a plan. (2024, August 9). The Verge. https://www.theverge.com/2024/8/9/24216329/cybersecurity-clean-energy-biden-administration-priorities The Verge

CyberAv3ngers: The Iranian saboteurs hacking water and gas systems worldwide. (2025, April 15). WIRED. https://www.wired.com/story/cyberav3ngers-iran-hacking-water-and-gas-industrial-systems

Cybersecurity of critical infrastructure. (2019). ResearchGate. https:// www.researchgate .net/publication/335752979_Cybersecurity_of_Critical_Infrastructure

Cybersecurity in power grids: Challenges and opportunities. (2021). arXiv. https:/ /arxiv. org/abs/2105.00013arXiv

Cybersecurity for critical infrastructure. (2023). IEEE Eta Kappa Nu. https://hkn.ieee.org/wp- content/uploads/2023/05/TheBridge_Issue2_May2023.pdf

Cybersecurity challenges in critical infrastructure. (2025). SciTePress. https:// www.scitepress. org/Papers/2025/130915/130915.pdfSciTePress

Cybersecurity and critical infrastructure resilience in North America. (2025). Wilson Center. https:// mexicoelections.wilsoncenter.org/publication/cybersecurity-and-critical-infrastructure-resilience-north-america

Cybersecurity and resilience bill. (2025). Wikipedia. https://en.wikipedia.org/wiki/Cyber_Security_and_Resilience_Bill

Cybersecurity strategies for critical infrastructure: Defending national security and ensuring resilience. (2025). ResearchGate. https:/ /www. researchgate. Net /publication/390486753_cybersecurity_strategies_for_critical_infrastructure_defending_national_security_and_ensuring_resilience

Cybersecurity strategies leveraging neural networks for critical infrastructure protection. (2025). IECE. https://iece.org/article/abs/tnc.2025.737491

Cybersecurity: State of the art, challenges and future directions. (2023). ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2772918423000188ScienceDirect

Dragos. (2025). 2025 OT cybersecurity report: 8th annual year in review. https://www.dragos.com/ot- cybersecurity-year-in-review/Dragos Cyber Security

Everything you need to know about cyber threat intelligence in 2025. (2025). Cyble. https://cyble.com/knowledge-hub/cyber-threat-intelligence-2025/

Examining cybersecurity critical infrastructure regulations in the U.S. and EU. (2025). Polsinelli. https://www.polsinelli.com/publications/examining-cybersecurity-critical-infrastructure-regulations-in-the-u-s-and-eu

National cyber threat assessment 2025–2026. (2025). Cyber Centre Canada. https:// www.cyber. gc.ca/en/guidance/national-cyber-threat-assessment-2025-2026Canadian Centre for Cyber Security

Securing the nation's critical infrastructures: A guide for the 2021–2025 administration. (2022). ResearchGate. https://www.researchgate.net/publication/364187815_Securing_the_Nation%27s_Critical_Infrastructures_A_Guide_for_the_2021-2025_Administration

Securing critical infrastructure in the age of AI. (2025). CSET Georgetown. https://cset.georgetown. edu/ publication/securing-critical-infrastructure-in-the-age-of-ai/CSET

Threat of cyber-attacks on Whitehall 'is severe and advancing quickly', NAO says. (2025, January 29). The Guardian. https://www.theguardian.com/technology/2025/jan/29/cyber-attack-threat-uk-government-departments-whitehall-naoThe Guardian

Top cybersecurity threats [2025]. (2025). University of San Diego. https://onlinedegrees.sandiego.edu/top-cyber-security-threats/University of San Diego Online Degrees

Towards a framework for improving cyber security resilience of critical infrastructures. (2025). Taylor & Francis Online. https:// www.tandfonline. com/ doi/ full/10.1080/12460125.2025.2479546