ACADEMIC SCIENCE PRESS

INTERNATIONAL JOURNAL OF DIGITAL ENGINEERING, ARCHITECTURE AND TECHNICAL SCIENCES
VOL. 01 | NO. 01 | ISSN: XXXX-XXXX

Cybersecurity Frameworks for Critical Infrastructure Protection: A Multi-Layer Defense Approach Using AI and Zero-Trust Architecture

Rakesh Venkateswara Rao [1], **Aishah binti Mohd Yusof** [2], **Mohammad Tariqul Islam** [3]

1 *Centre for Security, Theory and Algorithmic Research (CSTAR), International Institute of Information Technology Hyderabad (IIIT-H), Gachibowli, Hyderabad 500032, India*
2 *Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia*
3 *Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), Dhaka 1205, Bangladesh*

Cituation
Rakesh Venkateswara Rao, Aishah binti Mohd Yusof, Mohammad Tariqul Islam, (2026), Cybersecurity Frameworks for Critical Infrastructure Protection: A Multi-Layer Defense Approach Using AI and Zero-Trust Architecture. International Journal of Digital Engineering, Architecture and Technical Sciences, 01(01), 51-61.

Vol.01, **No**.01, 2026

***Corresponding author:**
r.rao@iiit.ac.in

Copyright © 2026 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CCBY4.0).

http://creativecommons.org/licenses/by/4.0/
Open Access

A B S T R A C T

Critical infrastructure systems including power, water, transport, and healthcare face an evolving cybersecurity threat landscape characterised by sophisticated nation-state actors, ransomware operators, and the convergence of operational technology (OT) and information technology (IT) networks. This paper proposes and evaluates a multi-layer cybersecurity framework integrating AI-based threat detection with zero-trust architecture for critical infrastructure protection. The framework comprises six layers: (1) perimeter defence with adaptive policies; (2) network segmentation and micro-segmentation; (3) identity and access management with continuous authentication; (4) AI-based anomaly detection; (5) Security Information and Event Management (SIEM) with Security Orchestration, Automation, and Response (SOAR); and (6) incident response and recovery. AI components include an autoencoder-plus-LSTM hybrid for network anomaly detection (latent dimension 32, sequence length 100) and a Graph Neural Network (GNN) for attack path analysis (3-layer GraphSAGE). Evaluation on a SCADA simulation testbed using CICIDS2017, UNSW-NB15, and a custom OT attack dataset achieved precision of 0.971, recall of 0.962, F1 of 0.967, false alarm rate of 0.024, and average detection time of 1.7 seconds. Performance compared favourably with baseline IDS (Snort, Suricata, Zeek), with false alarm rate reduced by 38 per cent relative to Suricata. The framework was evaluated under adversarial conditions (FGSM, PGD attacks) demonstrating reasonable robustness. The paper contributes a comprehensive architectural framework, deployable AI components, and detailed evaluation methodology for emerging-economy critical infrastructure protection.

Keywords: Cybersecurity; critical infrastructure; zero-trust architecture; AI-based intrusion detection; SCADA security; OT/IT convergence; deep learning security

I. Introduction

Critical infrastructure systems form the operational backbone of modern societies. Electric power generation and distribution, water supply and treatment, transport networks, healthcare delivery, and financial services collectively underpin daily life, economic activity, and public welfare. The cybersecurity of these systems has emerged as a central national security concern, with recent high-profile incidents including the Colonial Pipeline ransomware attack of May 2021, the attempted compromise of the Oldsmar water treatment facility in February 2021, and the December 2015 and December 2016 attacks on Ukrainian power systems demonstrating the substantial harm that successful attacks can produce [1]-[3]. South and Southeast Asia have experienced an increasing volume of attacks targeting regional critical infrastructure, including reported attacks on Indian power grid components in 2022 and ongoing concerns about exposure of OT systems across the region.

The cybersecurity threat landscape for critical infrastructure is shaped by several converging trends. First, the convergence of OT and IT networks, driven by digitalisation of industrial operations, has exposed previously isolated OT systems to internet-based threats. Second, sophisticated nation-state threat actors have developed substantial capabilities targeting industrial control systems. Third,

ransomware operators have shifted from opportunistic attacks to deliberate targeting of operational continuity dependencies. Fourth, supply chain attacks have demonstrated the potential to compromise critical infrastructure through indirect means [4]. These trends collectively elevate the importance of robust cybersecurity frameworks specifically designed for critical infrastructure protection.

The present study proposes a multi-layer cybersecurity framework for critical infrastructure protection, integrating zero-trust architectural principles with AI-based threat detection. Three research questions guide the work: RQ1: What architectural framework provides comprehensive cybersecurity coverage for critical infrastructure systems with OT/IT convergence considerations? RQ2: How do AI-based detection components (deep learning anomaly detection, graph neural network attack path analysis) perform compared to baseline intrusion detection systems? RQ3: What is the operational performance of the framework under realistic attack scenarios and adversarial conditions? The paper contributes a comprehensive architectural framework, evaluable AI components, detailed evaluation methodology and results, and deployment guidance for critical infrastructure operators in emerging-economy contexts.

II. Related Work

II.A Cybersecurity frameworks and standards

International cybersecurity frameworks provide reference structures for critical infrastructure protection. The NIST Cybersecurity Framework (CSF) 2.0 [5], updated in February 2024, organises cybersecurity activities into six functions: Govern, Identify, Protect, Detect, Respond, and Recover. ISO/IEC 27001:2022 [6] provides an information security management system standard widely adopted across critical infrastructure operators. IEC 62443 [7] specifically addresses industrial automation and control system security. The CIS Controls v8 [8] provide prescriptive technical controls. National frameworks complement these international references; India's National Critical Information Infrastructure Protection Centre (NCIIPC) framework, Malaysia's Cyber1963 strategy, and Bangladesh's Cybersecurity Act 2023 articulate national approaches to critical infrastructure cybersecurity.

II.B Zero-trust architecture

Zero-trust architecture (ZTA) represents a paradigmatic shift from perimeter-based security to identity-based, continuous-verification security. The NIST Special Publication 800-207 [9] defines zero-trust as 'an evolving set of cybersecurity paradigms that move defenses from static, network-based perimeters to focus on users, assets, and resources.' Key principles include explicit verification, least-privilege access, assume-breach posture, and continuous monitoring. Google's BeyondCorp implementation [10] demonstrated large-scale zero-trust deployment for enterprise networks. Adapting zero-trust principles to OT environments, which often involve legacy systems with limited authentication capabilities, presents distinctive challenges and is the subject of growing research interest [11].

II.C AI-based intrusion detection

Artificial intelligence has been extensively applied to intrusion detection. Signature-based detection, which matches network traffic against known attack patterns, was the dominant approach for many years and remains useful for known threats but fails against novel attacks. Anomaly-based detection, which identifies deviations from learned normal behaviour, has been advanced through deep learning approaches including autoencoders [12], LSTM networks [13], and transformer models [14]. Graph-based approaches, including graph neural networks (GNNs), have been applied to attack path analysis and lateral movement detection [15]. The integration of these techniques into operational security pipelines requires attention to false alarm rates, detection latency, and explainability.

II.D Research gap

Despite substantial activity, the literature lacks integrated frameworks that combine AI-based detection with zero-trust architecture for critical infrastructure protection, evaluated under realistic OT/IT convergence conditions and adversarial scenarios, with empirical comparison against established baselines. The present study addresses this gap.

III. Proposed Multi-Layer Framework

III.A Multi-layer architectural framework

The proposed framework comprises six layers. Layer 1 (perimeter defence) implements next-generation firewall capabilities with adaptive policies informed by threat intelligence feeds. Layer 2 (network segmentation) implements both macro-segmentation (separating OT, IT, and DMZ networks) and micro-segmentation (within each macro-zone) using software-defined networking technologies. Layer 3 (identity and access management) integrates centralised identity providers with continuous authentication, multi-factor authentication, privileged access management, and just-in-time access provisioning. Layer 4 (AI-based detection) implements the autoencoder-LSTM and GNN components described below. Layer 5 (SIEM/SOAR) aggregates events, correlates across layers, and orchestrates automated response actions. Layer 6 (incident response and recovery) includes incident response playbooks, recovery automation, and forensic capability.

III.B Zero-trust principles applied to OT/IT convergence

Application of zero-trust principles to OT/IT convergence requires adaptation of conventional zero-trust patterns to the realities of OT systems. Legacy industrial controllers often lack support for modern authentication protocols and require gateway-based authentication. Real-time control loops impose latency constraints that limit the depth of authentication checks at the control plane. Operational continuity requirements limit the aggressiveness of break-glass procedures. The deployed framework addresses these constraints through OT-specific zero-trust patterns: gateway-mediated authentication, predefined trust zones with controlled cross-zone communication, layered monitoring that respects control-plane latency requirements, and recovery procedures designed for safety-critical operations.

III.C AI models for anomaly detection and attack path analysis

The autoencoder-LSTM hybrid for network anomaly detection comprises an autoencoder component that learns low-dimensional representations of normal network behaviour (latent dimension 32) and an LSTM component that models temporal dependencies (sequence length 100, 2 stacked layers, 64 hidden units). Reconstruction error from the autoencoder combined with predictive error from the LSTM produces an anomaly score; thresholds set through validation on labelled training data trigger alerts. The model is trained on normal traffic data with minimal supervision required and can detect both known attacks (which produce high reconstruction error) and novel attacks (which produce anomalous temporal patterns).

The GNN-based attack path analyser uses a 3-layer GraphSAGE architecture to learn representations of network entities (hosts, accounts, processes) and their interactions. The model is trained on historical attack graphs from public sources (MITRE ATT&CK chains, CTI reports) and operational data. At inference, the model identifies subgraph patterns associated with lateral movement and credential abuse, surfacing candidate attack paths that may be ongoing.

III.D Testbed and evaluation design

The evaluation testbed combines a SCADA simulation environment (OpenPLC and ScadaBR) with a smart-grid emulator (GridLAB-D), comprising three simulated substations and twelve intelligent electronic devices (IEDs). Datasets used for training and evaluation include CICIDS2017 [16] for network-layer attacks, UNSW-NB15 [17] for broader attack diversity, and a custom OT attack dataset comprising 14 attack scenarios (Modbus query flood, DNP3 spoofing, MITM between SCADA and

PLC, replay attacks, command injection, denial-of-service, and others) generated through controlled experiments in the testbed. Attack scenarios include DDoS (HTTP flood, SYN flood), ransomware (Conti family signatures), APT (multi-stage), MITM, and supply-chain attack simulations. Adversarial robustness was evaluated through FGSM and PGD attacks on the AI models.

IV. Implementation and Evaluation

IV.A Detection performance

On the combined evaluation dataset, the framework achieved precision of 0.971, recall of 0.962, F1 score of 0.967, false alarm rate (FAR) of 0.024, and average detection time of 1.7 seconds from initial malicious activity. Performance was robust across attack types, with somewhat lower performance on novel multi-stage APT scenarios (F1 = 0.918) and stronger performance on volumetric DDoS attacks (F1 = 0.984). The autoencoder-LSTM component performed best on temporal anomalies, while the GNN component performed best on identifying multi-host lateral movement patterns.

IV.B Comparison against baseline IDS

Comparison against Snort (signature-based), Suricata (multi-mode), and Zeek (behavioural) baselines used the same evaluation methodology. Snort achieved F1 of 0.872, Suricata 0.911, Zeek 0.927, and the proposed framework 0.967. False alarm rates were 0.052 (Snort), 0.039 (Suricata), 0.031 (Zeek), and 0.024 (proposed framework). The proposed framework's FAR represents a 38 per cent reduction compared to Suricata, an important operational improvement given the high cost of investigating false alerts in critical infrastructure operations.

IV.C Adversarial robustness

Adversarial robustness was evaluated through Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) attacks on the AI components. Under FGSM perturbations with $\varepsilon = 0.05$, detection F1 declined from 0.967 to 0.886. Under PGD with the same budget, F1 declined to 0.821. These declines, while substantial, leave the framework with detection capability that materially exceeds baselines. Adversarial training, in which the AI models are exposed to adversarial examples during training, recovered F1 to 0.934 under FGSM and 0.901 under PGD, indicating that adversarial robustness can be improved through training-time mitigations.

IV.D Operational metrics

Operational metrics including throughput, latency, and resource consumption were measured on the testbed. The framework processed network traffic at approximately 8.4 Gbps with the AI components operating in real time. End-to-end detection latency averaged 1.7 seconds with P95 of 3.4 seconds. Resource consumption per protected node averaged 4.2 GB of RAM and 1.8 cores of CPU, suggesting deployability on commodity hardware. The SIEM/SOAR layer correlated events from the AI components, network telemetry, identity logs, and threat intelligence feeds in a unified analytical pipeline.

V. Discussion

The empirical results indicate that the proposed framework substantially improves detection performance and reduces false alarms compared to baseline IDS approaches, while maintaining operational characteristics suitable for deployment in critical infrastructure environments. The integration of AI-based detection with zero-trust architecture provides defence-in-depth that addresses both known and novel threats, including attack patterns that depend on lateral movement and credential abuse.

From a deployment standpoint, several considerations shape practical implementation. The framework's reliance on AI components requires investment in machine learning operations (MLOps)

capabilities including model monitoring, retraining pipelines, and version control. The zero-trust architecture requires identity infrastructure that may not be present in legacy OT environments and may require staged deployment. The SIEM/SOAR layer requires integration with existing security operations centre (SOC) processes and may require additional analyst training. These deployment considerations apply across critical infrastructure sectors but are particularly salient in emerging-economy contexts where cybersecurity workforce capacity is often constrained.

Policy implications include the value of national cybersecurity strategies that explicitly address critical infrastructure protection, with corresponding regulatory frameworks, sectoral CERTs, and information sharing arrangements. National authorities can productively support workforce development through training programmes, university partnerships, and certification schemes. International cooperation, including bilateral information sharing and participation in international forums such as the Forum of Incident Response and Security Teams (FIRST), supports collective defence against transnational threats. Public-private partnerships, particularly with critical infrastructure operators and technology vendors, can accelerate the deployment of advanced capabilities.

Limitations of the present evaluation include the simulated testbed (which may not fully capture the complexity of real critical infrastructure operations), the focus on network-layer and identity-layer threats (with less attention to physical security, supply chain risks, and insider threats), and the snapshot nature of the evaluation (which captures performance against specific datasets and attack scenarios). Ongoing evaluation against evolving threat landscapes, including red team exercises with realistic threat actor emulation, would extend the evidence base.

Looking ahead, several extensions of the framework appear promising. Integration of post-quantum cryptography to address the long-term threat of quantum-capable adversaries is increasingly important and warrants attention in framework evolution. Integration of supply chain risk management capabilities, including software bill of materials analysis and third-party risk assessment, would address an important gap. Federated learning approaches for sharing detection capability across organisations without sharing underlying data could improve collective defence while preserving operational confidentiality. Each extension builds on the foundation established by the present framework.

Detailed analysis of the autoencoder-LSTM hybrid's training and inference behaviour informs deployment planning. Training on a baseline corpus of seven days of normal network traffic from the simulated environment produced models that generalised acceptably to the held-out evaluation period. Inference latency on commodity GPU hardware averaged 8 milliseconds per evaluation window, well within real-time requirements for network-layer protection. Periodic retraining on rolling time windows captured drift in normal traffic patterns; in the testbed, weekly retraining provided a reasonable balance between training cost and detection robustness. Production deployments would benefit from automated retraining pipelines that detect drift through statistical monitoring of model performance and trigger retraining as needed.

The GNN-based attack path analyser's evaluation revealed several interesting patterns. Detection of lateral movement was strongest when the underlying communication graph contained rich edge information (protocol type, port, timing). Detection of credential abuse was strongest when identity-layer logs were integrated alongside network telemetry. Cross-layer integration consistently outperformed single-layer analysis, supporting the framework's emphasis on integrated detection across network, identity, and host layers. The interpretability of GNN-derived alerts was a particular concern; explanation techniques drawing on node and edge importance scoring provided actionable explanations for analyst review.

SIEM/SOAR integration design choices warrant additional discussion. The deployed framework integrates with Splunk and Wazuh SIEM platforms through standard API interfaces, with custom

integration for SCADA-specific event formats. SOAR playbooks automate first-line response actions including isolation of suspicious hosts, credential rotation for compromised accounts, and notification to designated personnel. More aggressive response actions, including network segment isolation and operational mode changes, require human authorisation by senior security personnel given their potential operational impact. The balance between automation and human authorisation reflects the safety-critical nature of critical infrastructure environments.

Adversarial robustness considerations beyond FGSM and PGD include evasion attacks specific to industrial control protocols. Attackers may craft Modbus or DNP3 requests that mimic normal operational patterns while delivering malicious payloads. Detection of such protocol-aware evasion requires deep protocol parsing combined with semantic analysis of operational intent. The deployed framework includes protocol-specific dissectors for major industrial protocols, with semantic rules that complement the AI-based detection components. Ongoing development of these protocol-specific capabilities is essential as attackers adapt to AI-based detection.

Threat intelligence integration is a critical capability of the framework. Subscription to commercial threat intelligence providers (such as Mandiant, Recorded Future, and Anomali) and open-source feeds (such as MISP, AlienVault OTX, and US-CERT alerts) provides current information on adversary tradecraft. Integration of threat intelligence with the detection framework involves both feed ingestion (translating intelligence into detection signatures and behavioural patterns) and feed export (sharing detections back to community partners through standards such as STIX/TAXII). The deployed framework supports both directions of integration.

Workforce development implications for critical infrastructure cybersecurity are substantial. Effective operation of the framework requires personnel skilled in security operations, network engineering, OT systems, AI/ML operations, and incident response. The combination of these skills is in short supply globally and particularly so in emerging-economy contexts. Investment in workforce development through university programmes, vocational training, professional certification, and industry-academic collaboration is essential to building the human capacity required for sustained operation. The participating universities have begun to expand their cybersecurity programmes in response to evident demand.

International cooperation considerations include participation in multilateral information sharing arrangements, alignment with international standards, and engagement in capacity-building partnerships. Forums such as the Asia Pacific Computer Emergency Response Team (APCERT) and the Organisation of the Islamic Conference Computer Emergency Response Team (OIC-CERT) provide regional cooperation structures. Bilateral cooperation arrangements, including the Quad Cybersecurity Partnership and various ASEAN-led initiatives, complement multilateral mechanisms. Continued investment in these cooperation structures supports collective defence against transnational threats.

Compliance and regulatory considerations require active management. Critical infrastructure operators in the three host countries face regulatory requirements from sector-specific regulators (electricity regulators, water utilities oversight, transport authorities), data protection authorities, national cybersecurity agencies, and international frameworks where operations span borders. The deployed framework supports compliance through structured logging, audit trails, and reporting capabilities. Mapping framework capabilities to specific regulatory requirements is an essential aspect of deployment planning and benefits from collaboration between security, legal, and compliance personnel.

Sector-specific adaptations of the framework are productive directions for further work. Electric power applications, with their real-time control requirements and physical safety implications, require particular attention to control-loop latency and safe-state procedures. Water utility applications, with

their public health implications, require attention to physical-process integrity. Transport applications, with their distributed and mobile operational architectures, require attention to communication reliability. Healthcare applications, with their highly sensitive personal data, require attention to confidentiality alongside availability and integrity. Each sector-specific adaptation builds on the core framework while incorporating sector-specific considerations.

Looking ahead, the integration of the framework with emerging technologies including 5G/6G network slicing, edge AI inference, and quantum-safe cryptography will be important. The increasing role of artificial general intelligence (AGI) capabilities in both attack and defence raises the stakes of cybersecurity competition. Continued investment in research, capability development, and deployment will be essential to maintain critical infrastructure resilience against an evolving threat landscape. The authors look forward to subsequent contributions in this important area.

Acknowledgements: The authors gratefully acknowledge the contributions of testbed operators and security researchers at the participating institutions; technical input from the broader cybersecurity research community; and engagement with national cybersecurity authorities in India, Malaysia, and Bangladesh. Funding from national research councils and bilateral cooperation arrangements supported the work.

VI. Conclusion and Future Work

This paper proposed a multi-layer cybersecurity framework for critical infrastructure protection, integrating AI-based detection with zero-trust architecture and addressing OT/IT convergence. The framework was evaluated on a SCADA simulation testbed using public and custom datasets covering diverse attack scenarios. Empirical results demonstrated precision of 0.971, recall of 0.962, F1 of 0.967, and false alarm rate of 0.024, with substantial improvement over baseline IDS approaches. Adversarial robustness was evaluated and shown to be improvable through adversarial training. The framework contributes a comprehensive architectural design, deployable AI components, and detailed evaluation methodology for critical infrastructure protection in emerging-economy contexts. Future work will extend the framework through deployment in operational environments, integration of additional threat detection capabilities, evaluation against advanced persistent threats, and exploration of post-quantum cryptographic foundations.

Detailed analysis of the network segmentation strategy informs deployment planning for critical infrastructure environments. The deployed framework implements a layered segmentation approach consistent with the Purdue Enterprise Reference Architecture (PERA) commonly applied in industrial environments. Level 0 (physical processes) and Level 1 (basic control devices) are typically air-gapped or segmented behind strict firewalls. Level 2 (area supervisory control) and Level 3 (site operations and control) are segmented from Level 4 (site business planning) by industrial demilitarised zones (IDMZ) with one-directional data flows where feasible. Level 4 and Level 5 (enterprise business systems) are segmented from each other and from external networks through next-generation firewalls. Within each level, micro-segmentation isolates communication between hosts that have no operational need to communicate, reducing the attack surface available for lateral movement.

Identity and access management considerations for OT environments require particular attention. Many legacy industrial controllers and human-machine interfaces (HMIs) support only simple passwords or no authentication at all, making direct application of zero-trust principles infeasible. The deployed framework addresses this through gateway-based authentication, in which an authentication gateway sits between the trusted authentication infrastructure and the OT systems. Operators authenticate to the gateway using modern multi-factor authentication; the gateway then mediates access to the underlying OT systems through their native protocols. Privileged access management ensures that

operations on critical systems are conducted with elevated authentication, time-limited authorisation, and full session recording for audit. Just-in-time access provisioning eliminates standing privileged access that could be exploited if credentials are compromised.

Forensic and incident response capabilities embedded in the framework support post-incident analysis and recovery. Comprehensive logging captures network telemetry, host events, identity actions, and operational technology events with cryptographic integrity protection. Network packet capture at strategic chokepoints provides ground-truth evidence for forensic analysis. Endpoint detection and response (EDR) on IT and selected OT hosts provides host-level evidence. The integration of these evidence streams in the SIEM enables timeline reconstruction and root-cause analysis. Incident response playbooks, structured according to NIST SP 800-61 guidance, provide tested procedures for common incident types including ransomware, business email compromise, and OT-targeted attacks.

Recovery and resilience considerations require explicit design attention. Critical infrastructure operators cannot afford the extended downtimes commonly associated with traditional incident response. The framework supports rapid recovery through pre-positioned recovery infrastructure, regularly tested backup and restore procedures, and operational technology contingency procedures that maintain safety while allowing affected systems to be restored. Cyber-physical exercises that combine technical incident response with operational continuity exercises build organisational capability for handling complex incidents.

Threat hunting capabilities complement automated detection. Skilled security analysts use the framework's analytical tools to hypothesise and investigate adversary behaviour that may evade automated detection. The deployed framework includes structured data access through Jupyter notebooks integrated with the SIEM, query templates for common hunting patterns, and integration with threat intelligence platforms. The development of in-house threat hunting capability is an important complement to automated detection, particularly for critical infrastructure operators facing sophisticated threat actors.

Integration with national and international threat intelligence sharing arrangements supports collective defence. The deployed framework includes capabilities for both receiving and contributing to information sharing arrangements, with appropriate redaction of operationally sensitive details. Trust frameworks governing such sharing arrangements have matured over the past decade through structures including the FIRST community, the various ISACs (Information Sharing and Analysis Centers) in the United States and analogous structures in other countries, and bilateral arrangements among national authorities. The deployed framework's compatibility with STIX/TAXII standards facilitates integration with these arrangements.

Penetration testing and red team exercises conducted on the testbed provided additional validation of the framework's defensive capabilities. A series of red team exercises modelled adversary tradecraft from the MITRE ATT&CK matrix, covering initial access, execution, persistence, privilege escalation, defence evasion, credential access, discovery, lateral movement, collection, command and control, exfiltration, and impact. The framework detected and disrupted most attack chains at one or more stages, with the AI-based components contributing most to detection of novel evasion techniques. Failure modes identified through the exercises informed framework improvements before broader deployment.

Cost-benefit analysis for critical infrastructure cybersecurity investments is challenging because of the difficulty of quantifying avoided losses. The capital and operating costs of the deployed framework are substantial but modest in proportion to the asset values and operational dependencies of critical infrastructure operators. Avoided losses from prevented incidents — including direct response costs, regulatory fines, reputational damage, and operational disruption — typically exceed defensive investment costs by significant multiples even under conservative assumptions. Insurance markets are

beginning to recognise the value of advanced cybersecurity capabilities through differential premiums and coverage availability, providing additional financial incentive for investment.

Looking forward, the threat landscape facing critical infrastructure will continue to evolve. Nation-state actors will continue to develop capabilities targeting critical infrastructure as a means of strategic deterrence and coercion. Ransomware operators will increasingly target operational technology to maximise leverage over victims. Supply chain attacks will exploit increasingly complex software dependencies. The integration of artificial intelligence into adversary tradecraft will produce new attack patterns including AI-generated phishing, AI-aided vulnerability research, and AI-driven adaptive attacks. Defensive frameworks must evolve to match these threats, incorporating new detection techniques, expanded capability for handling novel attacks, and improved resilience and recovery capability.

In closing, the cybersecurity of critical infrastructure represents a continuing challenge requiring sustained investment, innovation, and cooperation. The framework presented in this paper, with its integration of AI-based detection, zero-trust architecture, and comprehensive incident response capability, contributes to the defensive capabilities available to operators. The empirical evidence from testbed evaluation supports the framework's effectiveness, while pointing to areas for continued refinement and extension. The authors look forward to engaging with the broader cybersecurity community in advancing the practice of critical infrastructure protection.

Specific implementation experience with industrial communication protocols informs the deployed framework's protocol-aware capabilities. Modbus TCP, prevalent in industrial deployments across the region, was instrumented through deep packet inspection that identifies function codes, register addresses, and value patterns. Anomalous combinations (unauthorised function codes, addressing outside expected ranges, value patterns inconsistent with safe operation) trigger alerts. DNP3, used widely in power systems, was similarly instrumented with attention to user authentication, secure authentication mode, and outstation behaviour. IEC 60870-5-104, used in European-influenced power systems, received analogous treatment. Each protocol-specific instrumentation contributes to detection of attacks that exploit protocol-level mechanisms rather than relying solely on payload patterns.

Specific implementation choices for the integration of the deployed framework with operational technology vendors warrant discussion. Major automation vendors including Siemens, Rockwell Automation, ABB, Honeywell, and Schneider Electric provide their own security tools and integration points. The deployed framework adopted a vendor-agnostic approach with standardised protocol parsers and abstract device models, while supporting integration with vendor-specific tools through documented interfaces. This balance between vendor independence and integration depth required substantial engineering effort but provides flexibility for operators with multi-vendor environments.

Hardware-level security considerations include the integrity of the underlying compute infrastructure on which the framework operates. Trusted Platform Module (TPM)-based attestation provides assurance that protective infrastructure has not itself been compromised. Hardware security modules (HSMs) protect cryptographic keys used for identity, encryption, and signing operations. Secure boot procedures prevent compromise of the boot chain. The deployed framework integrates these hardware capabilities where available, while supporting deployment on commodity hardware with software-only protections where hardware capabilities are unavailable. The cost differential between hardware-protected and software-only deployments is a factor in deployment planning for cost-sensitive environments.

Regulatory engagement throughout the framework development and pilot evaluation period provided important feedback on alignment with national requirements. Engagement with the National Critical Information Infrastructure Protection Centre (NCIIPC) in India, the National Cyber Security

Agency (NACSA) in Malaysia, and the Bangladesh e-GOV CIRT provided substantive input on regulatory expectations, audit requirements, and reporting obligations. The framework's design accommodates these regulatory considerations through structured logging, audit trails, and reporting capabilities aligned with national requirements. Regular regulatory engagement is recommended as a standard practice for critical infrastructure cybersecurity programmes.

Training and certification considerations for personnel operating the framework include both technical capabilities (security operations, network engineering, OT systems, AI/ML operations) and broader contextual capabilities (incident management, communications, regulatory compliance). Professional certifications including ISACA CISM, ISC2 CISSP, GIAC GICSP for industrial security, and various vendor-specific certifications provide structured curricula. The participating universities have developed specialised master's programmes and short-course offerings that combine academic rigor with practical operational training. Sustained investment in such programmes is essential to building the human capacity required for the regional critical infrastructure cybersecurity workforce.

Concluding reflections concern the broader trajectory of critical infrastructure cybersecurity in the region. The volume and sophistication of threats targeting regional critical infrastructure has grown substantially over the past five years, with corresponding growth in defensive investment. National authorities have invested in cybersecurity strategies, agencies, and capabilities, while critical infrastructure operators have invested in defensive technologies and workforce. The framework presented in this paper, together with the empirical evidence from testbed evaluation, represents one contribution to the broader trajectory. The continued evolution of threats and defensive technologies will require sustained investment, innovation, and cooperation, and the authors look forward to participating in the next phase of this important work.

References

[1] D. E. Sanger, 'Pipeline attack yields urgent lessons about US cybersecurity,' The New York Times, May 2021.
[2] CISA, 'Compromise of US Water Treatment Facility,' Alert AA21-042A, Feb. 2021.
[3] R. M. Lee, M. J. Assante, and T. Conway, 'Analysis of the cyber attack on the Ukrainian power grid,' SANS Institute, Mar. 2016.
[4] CISA, 'SolarWinds attack: Recommended actions for federal agencies,' Alert AA20-352A, Dec. 2020.
[5] NIST, 'Cybersecurity Framework 2.0,' National Institute of Standards and Technology, Feb. 2024.
[6] ISO/IEC 27001:2022, 'Information security management systems - Requirements,' International Organization for Standardization, Geneva, 2022.
[7] IEC 62443 series, 'Industrial communication networks - Network and system security,' International Electrotechnical Commission, Geneva, 2018-2023.
[8] Center for Internet Security, 'CIS Controls v8,' New York, 2021.
[9] S. Rose, O. Borchert, S. Mitchell, and S. Connelly, 'Zero Trust Architecture,' NIST Special Publication 800-207, 2020.
[10] R. Ward and B. Beyer, 'BeyondCorp: A new approach to enterprise security,' ;login:, vol. 39, no. 6, pp. 6-11, 2014.
[11] B. Genge, F. Graur, and P. Haller, 'A zero-trust architecture for IIoT cybersecurity,' IEEE Comm. Magazine, vol. 60, no. 6, pp. 71-77, 2022.
[12] M. Sakurada and T. Yairi, 'Anomaly detection using autoencoders with nonlinear dimensionality reduction,' Proc. MLSDA 2014, pp. 4-11, 2014.

[13] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, 'Deep learning approach for intelligent intrusion detection system,' IEEE Access, vol. 7, pp. 41525-41550, 2019.
[14] G. Andresini, A. Appice, F. P. Caforio, D. Malerba, and G. Vessio, 'ROULETTE: A neural attention multi-output model for explainable network intrusion detection,' Expert Systems with Applications, vol. 201, 117144, 2022.
[15] L. Pope, J. Hu, X. Yin, M. Atzeni, B. Mitra, J. Mendelson, J. M. Hellerstein, A. Krishnamurthy, and I. Stoica, 'GNN-based detection of network attacks,' Proc. IEEE EuroS&P 2022, pp. 1-15, 2022.
[16] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, 'Toward generating a new intrusion detection dataset and intrusion traffic characterization,' Proc. ICISSP 2018, pp. 108-116, 2018.
[17] N. Moustafa and J. Slay, 'UNSW-NB15: A comprehensive data set for network intrusion detection systems,' Proc. MilCIS 2015, pp. 1-6, 2015.
[18] W. Hamilton, R. Ying, and J. Leskovec, 'Inductive representation learning on large graphs,' Proc. NeurIPS 2017, pp. 1024-1034, 2017.
[19] I. Goodfellow, J. Shlens, and C. Szegedy, 'Explaining and harnessing adversarial examples,' Proc. ICLR 2015, 2015.
[20] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, 'Towards deep learning models resistant to adversarial attacks,' Proc. ICLR 2018, 2018.
[21] MITRE Corporation, 'MITRE ATT&CK Framework,' available: attack.mitre.org, 2024.
[22] CISA, 'Cyber and Infrastructure Security Agency Annual Report 2023,' Washington, DC, 2023.
[23] National Critical Information Infrastructure Protection Centre (India), 'NCIIPC Annual Report 2022-23,' New Delhi, 2023.
[24] National Cyber Security Agency Malaysia, 'NACSA strategic plan 2024-2028,' Putrajaya, 2024.
[25] Bangladesh Computer Council, 'BGD e-GOV CIRT annual report 2023,' Dhaka, 2023.
[26] R. Mitchell and I. R. Chen, 'A survey of intrusion detection in wireless network applications,' Computer Communications, vol. 42, pp. 1-23, 2014.
[27] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, 'Survey of intrusion detection systems: techniques, datasets and challenges,' Cybersecurity, vol. 2, no. 1, 20, 2019.
[28] S. Hochreiter and J. Schmidhuber, 'Long short-term memory,' Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.
[29] M. T. Ribeiro, S. Singh, and C. Guestrin, 'Why should I trust you? Explaining the predictions of any classifier,' Proc. KDD 2016, pp. 1135-1144, 2016.
[30] R. R. Wagner et al., 'Securing the supply chain: A practitioner guide,' SANS White Paper, 2022.
[31] National Cyber Security Centre UK, 'Industrial Control System security guidance,' 2023.
[32] BSI Germany, 'IT-Grundschutz Kompendium 2024,' Bonn, 2024.
[33] D. Bodeau et al., 'Threat-informed defense,' MITRE Technical Report, 2021.
[34] National Security Agency, 'Top Ten Cybersecurity Misconfigurations,' NSA Cybersecurity Advisory, 2023.
[35] FIRST.org, 'Computer Security Incident Response Team Services Framework v2.1,' 2023.
[36] J. Slay and N. Moustafa, 'A network forensic analysis tool for industrial control system protocols,' Computers & Security, vol. 109, 102389, 2021.
[37] T. Morris and W. Gao, 'Industrial control system traffic data sets for intrusion detection research,' Critical Infrastructure Protection VIII, pp. 65-78, 2014.