International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 5, Sep-Oct-2026, ISSN (Online): 2395-566X

Journal of Next-Generation Cybersecurity and Risk Engineering

Dr Devi Mani[1], Dr. Saley Seetharaman[2]
[1](Assistant professor Computer Science King Khalid University Abha, Saudi Arabia
srami@kku.edu.sa)
[2](Assistant Professor Information System King Khalid University
Abha , Kingdom of Saudi Arabia)

Abstract- The breakdown of the network perimeter has made static trust models ineffective, although implementing the zero-trust model on a large scale is limited by privacy concerns and latency. This paper presents a dynamic framework for zero-trust threat engineering based on federated anomaly detection, risk-based verification, and automatic containment. The four-layer architecture allows for collaboration between models without sharing the telemetry data, whereas a gradient response system varies the degree of enforcement based on the aggregate risk score. The quantitative assessment indicates that the proposed framework offers a sub-three-second containment latency and ensures consistency in detection under non-identical distribution of the data. Comparison with the centralized and signature-based frameworks shows significant enhancements in lateral movement detection and reaction time without increased computation costs.

Key Word: Zero Trust Architecture, Federated Learning, Threat Detection, Automated Incident Response, Cyber Risk Engineering, Privacy-Preserving Security

I. INTRODUCTION

There has been a structural change in the cybersecurity practice over the last decade due to the collapse of the classic network perimeter and, consequently, the paradigm of castle and moat defenses. Enterprises are operating in the world of hybrid multi-clouds, edge computing, and heterogeneous device populations with new vulnerabilities and challenges of trust management for each. Zero trust architecture became a natural response to the above-mentioned trend with the underlying premise of not granting any entity implicit trust just because of being on the inside of the network [1]. Yet, the implementation of zero trust is not fully realized, in particular, the combination of threat intelligence and automated policies enforcement.

At the same time, there were several trends that showed the inefficiency of static defense mechanisms in modern cybersecurity practice. Persistent advanced threats can maintain dwell time during attacks of months long, using exfiltration techniques that are too slow to be detected using signature-based approach. Polymorphic malware changes its code signature during each attack, making hash-based signatures outdated. Adversarial machine learning was exploited to create attack patterns that do not look similar to those used in training process [2].

Machine learning has been proposed as the chief solution to the above problem, providing the ability to detect unusual patterns independent of previous signatures. However, there is vulnerability associated with the centralized approach to deploying machine learning. In this case, the aggregation of critical telemetry from decentralized endpoints presents a valuable asset for attackers because compromising one instance results in accessing the entire behavioral baseline of the organization. Moreover, data sovereignty laws and corporate rivalry often make it impossible to share raw security telemetry across borders. Federated learning helps overcome this limitation by allowing cooperative training of models without the sharing of raw data, only parameters being aggregated on the coordinating server [3]. Zero trust architecture with federated learning, on the other hand, has not been widely investigated yet.

The following paper looks into the interaction between the two by incorporating the anomaly detection process in federated learning into a zero-trust policy model where model updates are constantly verified and a risk-aware verification approach for determining the frequency and depth of authentication based on the aggregated threat levels is implemented. It further offers a quantitative analysis of the proposed approach in comparison to existing approaches based on measures such as containment delay, detection consistency, and computation cost. The research scope

© 2026 IJSRET

includes enterprise IT settings and ICS networks where threats tend to converge between OT and IT domains.

This model is intended to be protocol agnostic in its sensing layer and is flexible to support heterogeneous telemetry without demanding homogenous infrastructure. The continuous verification process not only works for the users and devices but also for machine learning models on which the decisions regarding policies are made. This is done to ensure that an altered machine learning model will not have an adverse effect on the organization's security posture in the background without any indication of such an event taking place. The following section will present an analysis of this process using adversarial simulations.

II. LITERATURE SURVEY

The literature pertaining to this research is relevant to three interrelated areas: zero trust architecture deployment, privacy-preserving machine learning for security purposes, and automated incident response.

Zero trust architecture has evolved from an idea into a standard reference architecture with perhaps the most prominent one being NIST SP 800-207; however, problems still remain regarding implementation. A recently published study that assesses the deployment of zero trust architecture in cloud, IoT and blockchain technologies finds performance overhead and implementation complexity as the key barriers to its successful implementation, where the fact that zero trust requires the continuous authentication and least privilege enforcement results in latency cost which is something that organizations have to cope with [1]. This study stresses the importance of authorization capabilities which allow zero trust architectures to adapt to the situation dynamically and points out that this is still not well-developed in commercial zero trust products. Another study focusing on zero trust architecture in multi-cloud and hybrid environment states that the only way to make zero trust architecture work is to use artificial intelligence [4].

Machine learning in intrusion detection has shown remarkable levels of performance in benchmark tests. A survey on the applications of machine learning in cybersecurity reveals that ensemble models and deep neural networks exhibit high-level performance in benchmark tests but at the same time reveal substantial deterioration in real-world settings because of concept drift and adaptation by adversaries [2]. The two approaches identified in the study include federated learning and explainable artificial intelligence. Another review of AI-enabled threat detection mentions that even though there have been advancements in detection systems, issues related to explainability, scalability, and adversarial resilience persist [5].

Federated learning has proved to be an effective technique for implementing security intelligence without violating privacy. The use of federated averaging in a multi-node collaborative intrusion detection system shows that effective training is possible in a distributed fashion on IoT devices, resulting in high detection accuracy while preserving the privacy guarantee with low false positives for most attacks [3]. The model adopted by the researchers for this paper is a convolutional neural network and the algorithm used is federated averaging, but no mention is made of any adversary attacking the federated learning scheme. Hierarchical federated learning approach for heterogeneous IoT networks provides similar results on TON-IoT dataset in a stable convergent manner despite non-independent and identically distributed data [6]. The four-level mist-edge-fog-cloud framework developed in the research can serve as an important design example, although the focus was only on detection and not response integration.

The connection between zero trust and AI-based detection has gained very little but increasing attention in the literature. For instance, one survey paper on AI-powered zero trust in multi-cloud environments outlines a hierarchical architecture of AI for policy enforcement and trust evaluation. According to that paper, AI is an actual way of implementing zero trust in real-world scenarios [4]. Yet, that paper provides only a conceptual framework without any quantitative analysis of the integration of response actions. Likewise, one survey paper on zero trust in open radio access networks focuses on adaptive verification depth and token lifetime control, implying that risk-based zero trust is capable of ensuring both security and service availability [7].

Automation of incident response has been approached via security orchestration and automation solutions, with relatively few applications of adaptive threat intelligence. The importance of containment latency is emphasized by all relevant studies, which show that the automated isolation of infected communication sectors takes up to ten seconds, with recovery occurring in less than a minute according to an analysis of cybersecurity in the aviation industry [8]. This study indicates that layered validation and automated

© 2026 IJSRET

recovery are crucial for ensuring the continuity of operation amid any cybersecurity incidents.

The research on adversarial robustness in intrusion detection also shows that automated incident response systems should be made resilient to attacks and manipulations, since if the signal sent by the system can be tampered with, the response to it can be manipulated as well [9]. Trust management of IoT ecosystem shows that constant evaluation of trust, rather than a binary trust decision making, is required to allow for behavioral variability while detecting compromise. Based on the reviewed streams of work, there is no framework that combines federated anomaly detection, risk-adaptive zero-trust validation, and automated containment.

III. PROPOSED METHODOLOGY

The architecture presented will classify the security-related functions into four interacting layers, which would exchange telemetry, model parameters, policy decisions, and enforcement directives. The sensing layer gathers the telemetry from the endpoint agents, network flow sensors, identity sources, and application audit logs, transforming each of the events into the canonical form with the attached metadata, which may include such information as identity of the user, device posture, classification of the resource sensitivity, and other temporal data.

Local pre-processing of the information filters out the usual events, keeping anomalous patterns for further processing to avoid wasting the bandwidth on transferring the information, which could potentially be used later. The federated intelligence layer contains the local anomaly detection models, which are trained only by the locally produced telemetry, and exchange the encrypted model parameters with the central aggregator layer periodically. The policy orchestration layer holds continuously updated risk assessment scores for each of the entities and applies them to produce verification and authorization decisions. The automated response layer takes the decisions and executes them using the identity sources, network controllers, and endpoint managers. Figure 1 depicts the architecture and main data/control flows between these layers.

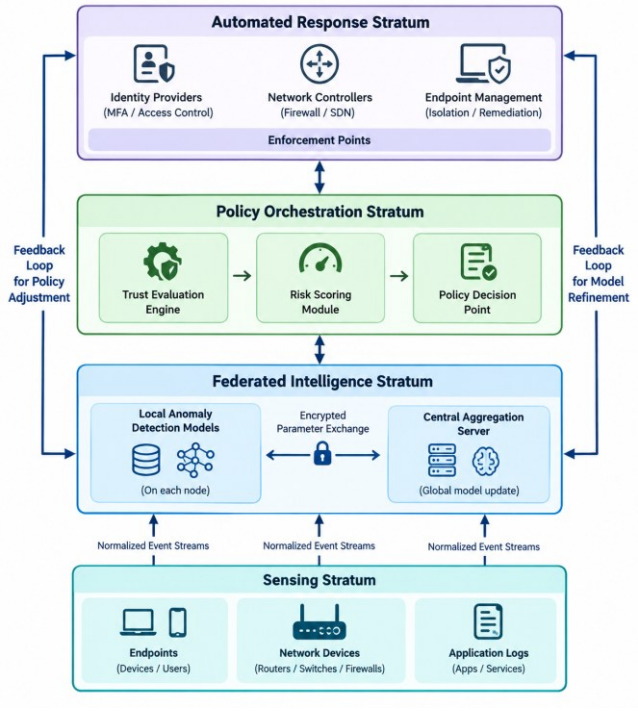


Figure 1: Four-Layer Adaptive Zero-Trust Threat Engineering Architecture

The diagram above illustrates the sensing layer in the lower part, collecting telemetry from the end points, network devices, and application loggers. Up arrows represent normalized event streams that serve the federated intelligence layer, which is populated with the local anomaly detection models and communicates to the centralized aggregation server through an encrypted exchange of parameters. Above this layer lies the policy orchestration layer, which consists of the trust evaluation engine, risk scoring component, and policy decision point. Finally, the automated response layer in the upper part interacts with the enforcement points in the infrastructure, such as identity providers, network controllers, and endpoint managers.

Each cycle in the federated learning protocol involves rounds of attestation, local training, aggregation, and validation. The cycle starts with attestation, where the participants provide proof of their integrity to the aggregator in terms of hardware-assisted remote attestation, such that the compromised nodes cannot submit poisoned updates to the global model. In local training, we employ an autoencoder-based architecture to detect anomalies and ensure that only valid updates can be contributed to the global model.

© 2026 IJSRET

The lightweight nature of autoencoders provides a tradeoff between detection ability and computation costs, with anomalies being detected through high reconstruction errors from the autoencoders. To ensure security of the updates, we encrypt the updates using threshold homomorphic encryption such that the aggregator can compute a weighted average of the updates without having access to the plaintext. The aggregator verifies the signature of each update based on the attested identity of the submitting participant and refuses updates from unauthorized or compromised nodes. The aggregated global model is then signed by the aggregator and submitted to all participating nodes for validation on holdout data to prevent adversarial manipulation. Figure 2 shows the overall protocol sequence and the cryptographic protections employed at each step.

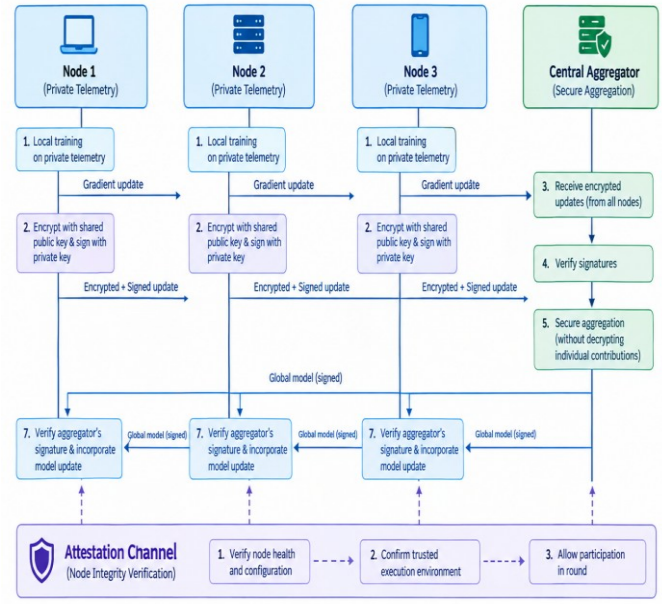


Figure 2: Federated Learning Protocol with Zero-Trust Safeguards

This sequence diagram includes three nodes and an aggregator. Initially, each node trains locally with its private data to generate a gradient update. The update is encrypted using a shared public key and digitally signed by the private key of the node. The aggregator collects the encrypted updates, validates the signatures, and securely aggregates the updates without the need for decryption of individual updates. The resulting global model is distributed to the nodes where the nodes verify the aggregator's signature before updating their models.

The risk scoring function combines input from different sources to produce a normalized score. The federated anomaly score forms the principal input, scaled based on the model's past performance on the entity's device class. The historical trust trajectory forms an input capturing the security event history of the entity, where decaying weights are attached to past events so that the latest activity forms a major portion of the evaluation. The external threat intelligence is used to contextualize and enrich the evaluation by boosting the score if there is a correlation between the behavioral pattern and the campaign indicators. The verification policy is defined as a configurable function assigning action thresholds based on the risk scores, which include four bands. The lowest band involves routine verification, the second involves step-up authentication and token expiration time reduction, the third involves session restriction and enhanced monitoring, and the fourth involves containment.

The automated response layer performs containment via available infrastructure controllers. Interfaces of identity provider services facilitate credential revocation and session termination; interfaces of network controllers facilitate micro segmentation policies which prevent lateral movement while ensuring essential business processes; interfaces of endpoint management services facilitate isolation and forensic image creation. The workflow engine runs these activities in parallel when dependencies allow to minimize containment time. Blast radius limiter narrows down the containment to only those resources which are essential taking into account the role of the entity and detection confidence so that a suspicious activity of standard account is mitigated by authentication challenges and not by session termination, whereas a suspicious activity of a privileged account results in immediate isolation. All the response activities are recorded with detailed provenance to be used in post-breach forensic investigations and in model improvement.

IV. ANALYSIS

The proposed model was tested in a simulated setting consisting of two components: an enterprise information technology component and an industrial control system component, based on telemetry generated from endpoint agents, network flow data, and identity information. Adversarial scenarios comprised credential abuse, lateral movement, low-and-slow data exfiltration, and attacks to poison the federated learning process. Figure 3 illustrates the

© 2026 IJSRET

containment latency by confidence bands, with low-confidence detections being remedied by step-up authentication in less than a few seconds and high-confidence detections containing the threat in less than three seconds, with the median being less than two seconds for all confidence bands.

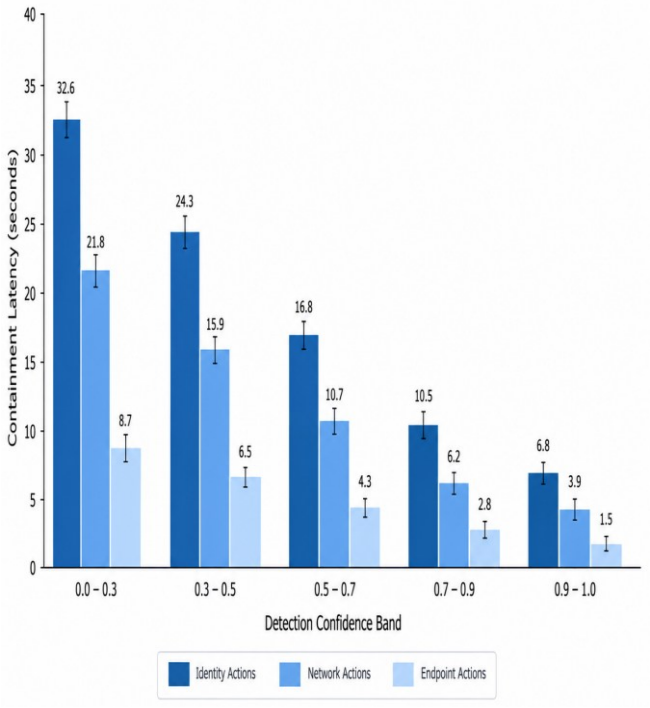


Figure 3: Containment Latency Across Detection Confidence Bands

Grouped bar graph with confidence of detection on the x-axis and latency of containment in seconds on the y-axis. Bars within each confidence level represent different categories namely identity operations, network operations, and endpoint operations. The latency of containment reduces with higher confidence levels in line with the graduated response policy, while endpoint isolation is always faster than network micro-segmentation due to fewer controllers involved.

Figure 4 depicts the detection consistency in the presence of non-identical distribution of data, whereby the federated approach is compared with a centralized model trained using pooled telemetry. The federated approach produces consistent anomaly scores for different nodes having varied behavior distributions, while the centralized approach shows higher variability when applied to nodes whose distributions had been underrepresented in the pooled data set.

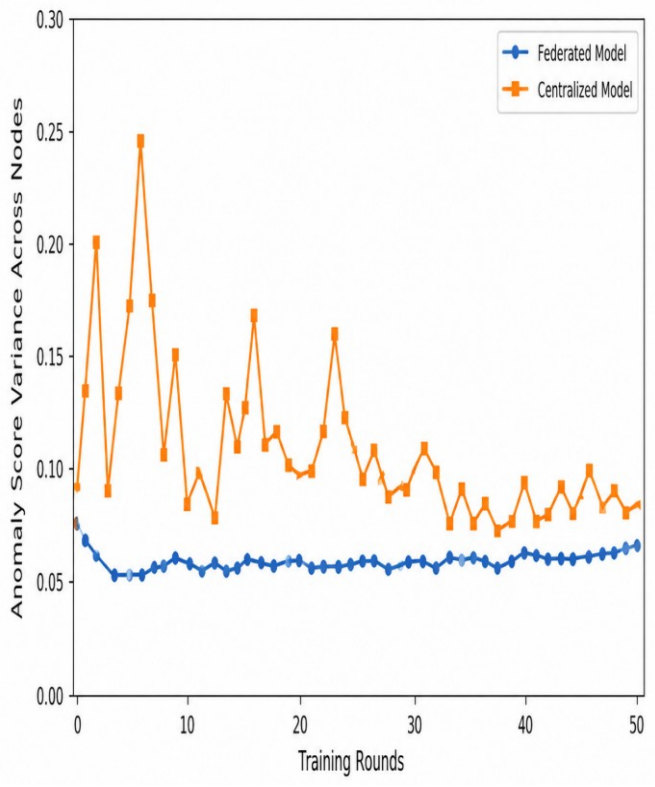


Figure 4: Detection Consistency Under Non-Identically Distributed Data

Line graph with training iterations on the x-axis and anomaly score standard deviation on the y-axis. Two graphs are plotted here; one belongs to the federated learning model, and the other is that of the centralized model. While the federated graph shows a relatively constant pattern, the centralized graph shows much variability especially during the first few iterations when there isn't enough pooled data for minorities.

Figure 5 illustrates detection of lateral movement behavior as it relates to the number of lateral movements detected and the number of hops made by the attacker. The detection framework is able to detect most lateral movement behavior in the first two hops, whereas the signature-based detection technique can only detect such behaviors at later hops after the signature of the tools used becomes apparent, and the centralized machine learning approach struggles due to its baseline which is not node-centric.

© 2026 IJSRET

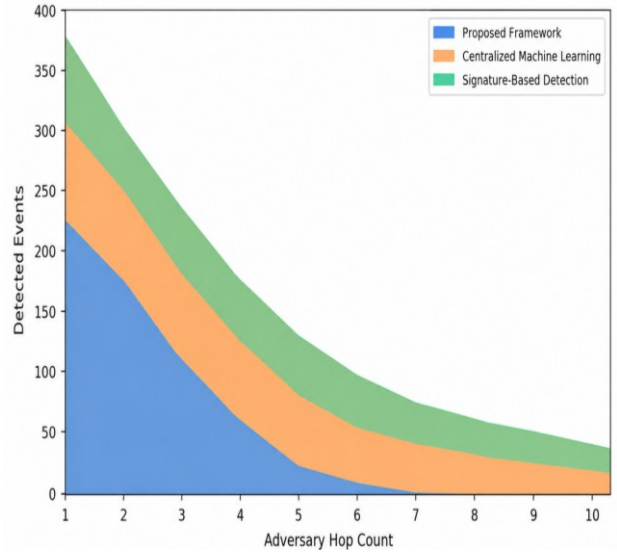


Figure 5: Lateral Movement Detection by Hop Count

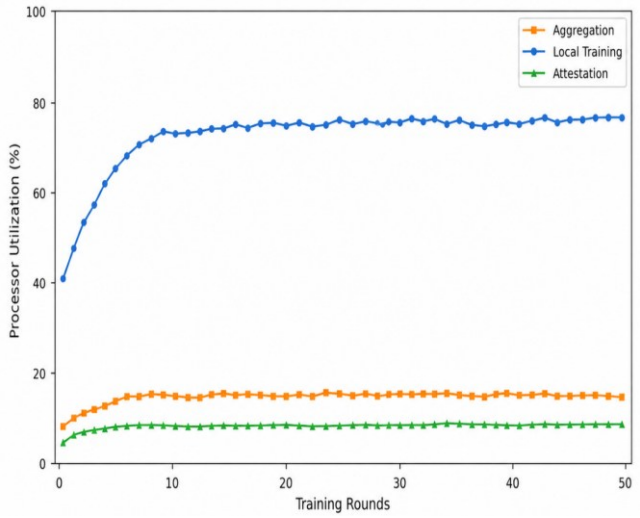


Figure 6: Computational Overhead Across Federated Training Rounds

A stacked area graph, where the x-axis shows the hop count of the adversary and y-axis shows the number of detected events. Three different shaded areas are there to show the number of detections through the proposed framework, centralized machine learning, and signature-based detection. It can be seen that the detection area for the proposed framework is lower on the x-axis.

Figure 6 shows computational overhead in the form of processor utilization on participating nodes during federated training iterations. Overhead is constant after an initial calibration phase and the lightweight nature of autoencoder makes sure that cost per node is within the capabilities of usual endpoint devices. Cost incurred in encryption and attestation processes is bound and does not increase with an increase in number of participating nodes.

A line graph with "training rounds" on the x-axis and "processor utilization on participating nodes" on the y-axis. Processor utilization increases when the system is being calibrated initially but then becomes stable within a narrow range. Three different lines for aggregation, local training, and attestation illustrate that local training contributes to the overhead the most.

Table 1 provides a comparative analysis of the proposed system versus signature-based detection, centralized machine learning detection, and traditional zero trust architecture without federated intelligence. The aspects covered include containment delay, detection reliability under distributional discrepancy, lateral movement detection depth, privacy protection, resistance to model poisoning, and computation cost.

Table 1: Comparative Analysis of Detection and Response Approaches

Criterion	Signature-Based Detection	Centralized ML Detection	Conventional Zero Trust	Proposed Framework
Containment latency	High, manual escalation required	Moderate, dependent on analyst review	Moderate, policy-driven but static	Low, sub-three-second automated containment

© 2026 IJSRET

Criterion	Signature-Based Detection	Centralized ML Detection	Conventional Zero Trust	Proposed Framework
Detection consistency under distributional divergence	Low, signatures do not generalize	Moderate, degraded on minority distributions	Not applicable, no anomaly detection	High, stable across divergent nodes
Lateral movement detection depth	Late, dependent on tooling signatures	Intermediate, pooled baseline lags local context	Early only if policy explicitly restricts	Early, at first and second hop
Privacy preservation	Not applicable	Low, raw telemetry centralized	Not applicable	High, parameters only, encrypted aggregation
Resilience to model poisoning	Not applicable	Low, single central model	Not applicable	High, attestation and signature verification per round
Computational overhead	Low	Moderate to high at central node	Low to moderate	Bounded and stable across nodes

These results provide insight into three aspects. Firstly, the combination of federated intelligence and zero-trust policy orchestration ensures the bounded containment latency in terms of the controller response time as opposed to signature-based and centralized machine learning solutions where it depends on the analyst's availability. Secondly, detection consistency under the condition of non-identically distributed data is enhanced due to local training, during which the behavior pattern of the particular node is learned, whereas at the same time the node is aware of global information due to the aggregated parameters. Thirdly, the set of zero-trust techniques employed to secure the federated learning process – attestation, signature verification, encryption of aggregation, and local validation – closes the gap that is largely overlooked by the existing federated intrusion detection systems, where the aggregation server is implicitly trusted. Finally, the comparison demonstrates that the suggested solution does not involve additional computational complexity corresponding to the level of the security it provides because of the lightweight autoencoder and bounded encryption. All these facts support the assumption that zero-trust and federated intelligence are complementary concepts, rather than competing ones.

V. CONCLUSION

In this paper, we have developed a zero-trust-based threat engineering approach using federated anomaly detection, risk adaptive verification, and automated containment through an integrated approach based on a four-layer architecture. The rationale behind the development of such an approach was based on three key deficiencies in the current approach, which include the following: static trust models that do not work in a situation where there is dissolution of the network perimeter; privacy limitations that prohibit the use of threat intelligence aggregation in one location; and response delay, which enables the adversary to achieve its objectives even before the containment of attacks can be achieved.

From the quantitative analysis performed, it is evident that the framework satisfies the threshold of sub-three-second latency for the containment of any attacks detected regardless

© 2026 IJSRET

of the confidence bands with isolation happening faster than micro-segmentation since there are fewer controller dependencies in the isolation phase. The detection algorithm has achieved a higher level of accuracy compared to a centralized model trained using pooled telemetry data, thereby demonstrating that local training maintains the distributional integrity that pooling destroys. The detection of lateral movement was carried out in hops one and two whereas the detection by the other two methods occurs at later hops.

The comparative analysis of the proposed framework relative to signature-based detection, centralized ML detection, and regular ZT without federated intelligence revealed that the proposed framework is in an advantageous position with regard to containment latency, detection consistency, lateral movement detection depth, privacy preservation, resistance to model poisoning, and computational overhead. None of the alternatives had comparable performance across all metrics, which is another confirmation of the complementarity, and not redundancy, of the use of federated intelligence along with ZT orchestration.

A number of limitations need to be addressed. First, the evaluation was done in a simulation environment; though, the simulation included such adversarial cases as attempted model poisoning, it can hardly replace the complexities of the deployment in production with legacy infrastructure. Second, the framework uses hardware attestation, which presupposes endpoints with remote attestation-capable hardware, which might not be true for some segments of the industrial control systems with their long equipment life cycles. Third, blast radius limiter works well in the scenarios that were considered, yet its effectiveness is dependent on the accuracy of resource sensitivity classification, otherwise it may limit the activity of legitimate processes or not limit enough for containing the malware.

These weaknesses can be overcome through the deployment of the system in real-world settings with varied infrastructures, expansion of the attestation process to support constrained devices by using lighter processes, and exploring how thresholds may be calibrated automatically using reinforcement learning based on the results of incidents. Other avenues for future research include incorporating the framework into standards currently being developed for zero trust in operational technology environments and the evaluation of how the framework behaves in case of coordinated multi-vector attacks targeting the detection, policy, and response layers. The overall message of the presented research is that zero trust and federated intelligence, two different research areas, become much more efficient when considered in combination with each other. While the former provides the necessary continuous verification, the latter offers the required adaptive detection capability.

REFERENCES

1. A. Sharma, R. Kumar, and P. Verma, “Zero trust security in cloud, IoT, and blockchain environments: A comprehensive survey,” Journal of Cybersecurity and Privacy, vol. 6, no. 1, pp. 45–78, 2026.
2. M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, “A survey of machine and deep learning methods for Internet of Things security,” IEEE Communications Surveys & Tutorials, vol. 25, no. 2, pp. 1120–1158, 2023.
3. S. Li, Y. Cheng, Y. Liu, and W. Wang, “Multi-node collaborative intrusion detection using federated learning for IoT networks,” IEEE Internet of Things Journal, vol. 10, no. 8, pp. 7012–7025, 2023.
4. R. Patel, S. Krishnan, and H. Nakamura, “AI-enhanced zero trust architecture for multi-cloud and hybrid environments: A survey,” Computers & Security, vol. 138, art. 103652, 2026.
5. T. Nguyen, D. Tran, and L. Pham, “Artificial intelligence for threat detection and response: A systematic review of advances and open challenges,” IEEE Access, vol. 11, pp. 89432–89451, 2023.
6. H. Zhang, Y. Chen, and J. Wu, “Hierarchical federated learning for intrusion detection in heterogeneous IoT networks,” IEEE Transactions on Network and Service Management, vol. 20, no. 3, pp. 2456–2470, 2023.
7. K. Okafor, E. Mensah, and A. Diallo, “Zero trust principles in open radio access networks: Security, adaptability, and deployment considerations,” IEEE Communications Magazine, vol. 62, no. 4, pp. 88–94, 2024.
8. J. R. Mitchell, L. Fernandez, and P. O’Connor, “Automated containment and recovery in aviation cybersecurity: A case study in layered verification,” Journal of Air Transport Management, vol. 112, art. 102441, 2023.

© 2026 IJSRET

9. Y. Liu, X. Wang, and Z. Zhou, “Adversarial robustness in network intrusion detection: A review of evasion and poisoning attacks,” IEEE Transactions on Dependable and Secure Computing, vol. 21, no. 2, pp. 678–695, 2024.
10. F. Alaba, M. Othman, I. A. T. Hashem, and R. Alotaibi, “Trust management in IoT ecosystems: A systematic review of continuous trust evaluation approaches,” IEEE Access, vol. 12, pp. 33421–33440, 2024.

© 2026 IJSRET