Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Bridging the Gap between Network Security and AI-Driven Threat Detection

Namboodiri Arun Mullamangalath Kesavan
Northwestern University, USA

ARTICLE INFO

Received: 02 Oct 2025

Revised: 04 Nov 2025

Accepted: 14 Nov 2025

ABSTRACT

Modern network architectures operating throughout distributed cloud environments and encrypted communication channels have rendered conventional perimeter-based safety mechanisms incapable of detecting advanced persistent threats. Artificial intelligence techniques employed by adversaries to craft polymorphic malware, automate reconnaissance activities, and cover command-and-control communications within legitimate protocol traffic require essential transformation of protective capabilities. Conventional signature-based and rule-driven detection systems are unable to evolve and identify behavioural anomalies across encrypted traffic flows, generating extensive blind spots that sophisticated attackers systematically exploit. The article discusses the strategic embedding of machine learning methodologies within network security architectures through robust data pipelines, real-time inference mechanisms, and continuous learning frameworks. Hybrid deep learning architectures, which combine convolutional neural networks with bidirectional long short-term memory components, have emerged as superior in capturing spatial features and temporal dependencies inherent in network telemetry streams. Implementation challenges include extreme class imbalance driven by rare malicious traffic samples, latency constraints necessitating millisecond-scale inference for inline enforcement, interpretability requirements enabling analyst comprehension of detection rationale, and adversarial attacks aimed at compromising training data integrity or crafting evasive inputs. Operational deployment requires comprehensive telemetry collection across heterogeneous sources, advanced feature engineering transforming raw packet data into statistical representations, and seamless integration with security orchestration systems. Augmented intelligence frameworks establishing bidirectional collaboration between automated detection systems and human analysts allow for continuous model refinement through labelled feedback loops and enable adaptive defense ecosystems capable of evolving alongside emerging threat landscapes.

Keywords: Artificial Intelligence, Network Security, Intrusion Detection, Machine Learning, Encrypted Traffic Analysis, Hybrid Cloud Architecture

Introduction

Cloud adoption, software-defined wide area networks, and distributed workforces dissolve conventional network perimeters and create a completely different threat landscape. Today's enterprises manage network architectures where organizational workloads operate across hybrid cloud environments, creating dispersed attack surfaces well beyond conventional data centre boundaries. The transition to hybrid cloud infrastructure introduces substantial complexity in security governance, as organizations must manage on-premises systems while integrating both public cloud services and private cloud resources [1]. Traffic flows across heterogeneous environments where distinguishing trusted from untrusted domains becomes increasingly complex; each enterprise network can handle traffic from multiple cloud service providers all at the same time. The architectural challenges entailed in hybrid cloud adoption involve not just technical difficulties in system integration but also fundamental issues of data sovereignty, compliance with varied jurisdictions, and a lack of standardized security frameworks across different cloud platforms [1].

Simultaneously, the proliferation of encrypted communications—while vital to privacy—builds opacity that obscures malicious activity from traditional inspection mechanisms. Current network telemetry reveals that encrypted traffic has become the dominant form of internet communications, with Transport Layer Security and encrypted Domain Name System queries becoming universal standards that simultaneously cloak user communications from inspection while masking potential threat indicators from security monitoring systems. Threat actors take advantage of this complexity by using AI-generated evasive payloads, automating vulnerability scanning, and hiding C2 communications within standard protocol traffic. Recent analyses show that these actors now use machine learning algorithms to develop polymorphic malware variants whose signature changes at unprecedented velocities, fundamentally outpacing traditional signature database update cycles that operate on extended refresh periods [2]. The introduction of adversarial machine learning techniques has provided threat actors with a method to systematically probe detection systems for their decision boundaries and craft evasive variants that exploit weaknesses in those systems while retaining functional payload integrity [2].

Traditional defence mechanisms that rely on predefined signatures and static ruleset evaluations prove inadequate in combating such adaptive threats. Signature-based intrusion detection systems show significant detection latencies for novel attack variants, allowing temporal windows where adversaries establish persistent access and exfiltrate sensitive data. Rule-based firewall configurations, though often effective against known threat patterns, consistently demonstrate considerable false negative rates when confronted with zero-day exploits or methodologies of attack that do not conform to documented threat intelligence. The integration of real-time malware detection capabilities requires computational architectures capable of processing high-velocity data streams while maintaining detection accuracy against adversarially designed evasion techniques [2]. Security operations centres mentioned that conventional monitoring tools provide substantial alert volumes, which, upon further investigation, are found to be mostly false positives triggered by legitimate administrative activities, known application behaviours, or misconfigured detection thresholds.

This article explores how behavioural analysis, powered by artificial intelligence, can transform network security from reactive pattern matching to predictive anomaly detection in a way that addresses fundamental architectural and operational challenges involved in this paradigm shift. By establishing baseline behavioural models that learn normal network communication patterns, statistical flow characteristics, and temporal access sequences, machine learning systems are able to find deviations indicating compromise even when specific attack signatures are unknown.

Limitations of Traditional Network Security Approaches

Conventional protection architectures are designed around layered protection strategies that include firewalls, intrusion detection and prevention systems, security information and event management systems, and endpoint protection mechanisms. These solutions were designed for static, perimeter-centric environments where network boundaries and asset locations remained relatively stable. The architectural basis of traditional security models assumes clear demarcation between internal trusted networks and external threat domains—an assumption fundamentally undermined by modern distributed computing paradigms. This model is increasingly ineffective with contemporary network architectures, as the cloud-native applications, microservices deployments, and hybrid infrastructure configurations remove any notion of a defensible perimeter. Intrusion detection methodologies have evolved through successive stages, starting with signature-based approaches matching known attack patterns, progressing to anomaly-based techniques that define deviations from established baselines, and more recently, hybrid frameworks attempting to combine both paradigms.

The intrinsic limitations of traditional approaches include reliance on signature databases that cannot anticipate novel attack vectors; the inability to analyze encrypted traffic without introducing considerable computational overhead and privacy concerns; and the generation of an excessive number of false positive alerts, which overwhelm security operations centers. Signature-based intrusion detection methodologies are fundamentally based on previously known threat patterns, with an inherent temporal lag between the onset of a particular attack and deployment of the corresponding defensive capability. Indeed, intrusion detection systems using signature matching methods prove very effective against cataloged threats for which patterns are known within their knowledge bases, but often demonstrate significant performance degradation when confronted with polymorphic malware variants or zero-day exploits without corresponding signature entries within their pattern databases. Taxonomy of intrusion detection approaches. In this regard, intrusion detection approaches based on signature systems have low false positive rates and are computationally efficient; however, the inability of these systems to detect new attacks is a fundamental weakness that cannot be improved by incremental enhancements of pattern databases [3]. Anomaly-based detection systems strive to address this shortcoming by establishing baseline models of normal behavior and flagging deviations; however, they face severe challenges related to defining appropriate thresholds that balance detection sensitivity against false positive generation.

Rule-based detection mechanisms require explicit definitions of malicious patterns that create blind spots for previously unseen attack methodologies. Defining detection rules requires exhaustive knowledge about attack vectors and their manifestations in network traffic, which is not available in the case of emerging threat categories. The challenge of encrypted traffic analysis is even more critical, as modern encryption protocols hinder deep packet inspection while simultaneously carrying malicious communications. The spread of encrypted communications over Transport Layer Security and similar protocols has relegated traditional payload inspection approaches to oblivion, leaving security systems to identify threats based exclusively on metadata and flow characteristics. Research into encrypted traffic classification illustrates that converting network flows into visual representations via temporal binning and packet size encoding allows the use of image recognition techniques to differentiate between categories of traffic without decryption [4]. This transforms sequential network flow data into two-dimensional matrices in which temporal progressions map to one axis and packet size distribution to the other, yielding a distinctive visual pattern that changes systematically across application types.

Real-time traffic inspection at contemporary network speeds creates scalability challenges that limit the comprehensiveness of monitoring capabilities. High-velocity networks operating at substantial throughput rates generate packet volumes that exceed the processing capacity of inline inspection systems, forcing organizations to implement selective monitoring strategies that introduce coverage gaps. The high computational intensity of traditional classification techniques, particularly those requiring feature extraction from individual packets or complex statistical analysis across flow sequences, constrains deployment feasibility in production environments handling substantial traffic volumes. Encrypted traffic classification by flow visualization techniques overcomes these scalability limitations by providing the ability to parallel process converted flow images through convolutional neural network architectures with classification latency compatible with real-time monitoring requirements [4].

These limitations call for a switch toward behavioral analysis systems that are capable of learning normal network patterns and identifying statistical deviations that indicate compromise, even within encrypted traffic flows. Machine learning processes offer the opportunity of moving past signature-based boundaries by recognizing patterns that represent anomalous behavior, as opposed to matching known attack signatures, consequently enabling the detection of new threats that do not depend upon prior threat intelligence.

Security Componen t	Design Basis	Key Limitations	Detection Impact
Signature-Based Detection	Known attack pattern matching	Requires prior threat knowledge; delayed signature updates	Effective for catalogued threats; fails against polymorphic malware and zero-day exploits
Rule-Based Firewalls	Predefined malicious behaviour rules	Cannot detect novel attacks; needs explicit threat definitions	Blind spots for unseen attacks; high false positives with broad rules
Encrypted Traffic Inspection	Deep packet payload analysis	Encryption prevents payload visibility; high computational cost	Limited to metadata analysis; raises privacy concerns
Real-Time Monitoring	Inline high-speed traffic inspection	Capacity constraints at multi-gigabit rates	Requires traffic sampling; creates coverage gaps

Table 1. Limitations of Traditional Network Security Mechanisms [3, 4].

AI-Driven Detection Methodologies and Technical Challenges

Artificial intelligence transforms threat detection from explicit rule matching into a function of statistical pattern understanding. Supervised learning methods train the models on a labelled dataset of network traffic samples, including malicious and benign samples, to allow the classification of new observations through pattern recognition algorithms trained to learn discrimination features from historical examples. Applications of machine learning to intrusion detection are really broad and span a variety of algorithmic approaches, from traditional classifiers to more complex deep learning architectures able to process sequential data of network traffic. Indeed, recent studies have demonstrated that hybrid convolutional neural networks with bidirectional long short-term memory obtain state-of-the-art results in intrusion detection tasks by embedding two layers for feature extraction. Convolutional layers extract the spatial features from representations of network traffic, aiming at discovering local patterns and correlations within sequences of packets, while the bidirectional long short-term memory components model temporal dependencies by processing network flows both in the forward and backward temporal directions [5]. This process of bidirectional processing enables the detection system to contextualize current network events within both preceding and subsequent traffic patterns, enhancing recognition of attack sequences that develop over greater intervals of time.

Unsupervised learning approaches identify unknown threats through clustering of similar behaviours and flag statistical outliers that do not align with established patterns. These are especially useful for finding new attack variants that do not have any representation in training datasets, since clustering algorithms will tend to group network flows due to inherent similarity measures without any prior labeling of malicious versus benign traffic. Deep learning architectures, especially long short-term memory networks and convolutional neural networks, capture temporal dependencies and spatial correlations within traffic flow data that may not be discernible by traditional methods. The architecture of deep learning models for intrusion detection should be carefully designed with respect to feature representation strategies, and research indicates that direct encoding of raw packet bytes or protocol fields as input tensors enables end-to-end learning without explicit feature engineering, although this increases computational requirements during training phases [5].

ML Technique	Core Components	Detection Strength	Key Considerations
Supervised Learning	Decision trees, support vector machines, ensembles	Classifies known attack patterns from labelled data	Needs comprehensive labelled datasets; quality-dependent performance
Unsupervised Learning	Clustering, outlier detection	Identifies novel threats without prior labels	Threshold tuning balances sensitivity and false positives
Hybrid CNN-BiLSTM	Convolutional layers + bidirectional LSTM	Captures spatial patterns and temporal dependencies	Higher training computation eliminates manual feature engineering
Ensemble Methods	Random forests, gradient boosting	Robust classification across diverse attacks	Superior accuracy with deployment-ready efficiency

Table 2. Machine Learning Techniques for Network Intrusion Detection [5]

While these are effective in controlled testbeds, applying them in production environments presents significant challenges. Malicious traffic accounts for only a small portion of the total amount of network activity, leading to a high class imbalance that skews model training towards benign classifications. The fact that operational networks predominantly contain normal traffic means their corresponding training datasets have malicious examples in notably smaller proportions compared to benign flows, making standard learning algorithms achieve high overall accuracy but fail to detect the minority attack class. Model inference needs to occur inside strict time bounds so that it can enable inline traffic enforcement, greatly constraining architectural choices and computational complexity. Security analysts need explainable outputs to comprehend detection rationale and allow them to integrate domain expertise, yet many high-performance models are inherently black boxes. In particular, the interpretability challenge is very serious with deep neural networks, wherein the detection decisions arise from complex nonlinear transformations across multiple hidden layers.

Adversarial machine learning further empowers attackers with the capabilities to poison training datasets or craft inputs designed to evade detection algorithms. Security threats against machine learning systems involve several attack vectors along a data processing pipeline, starting from training data collection down to model deployment and inference stages. Poisoning attacks aim at manipulating the training dataset by either injecting malicious samples or corrupting existing benign samples. This results in models learning incorrect decision boundaries, enabling subsequent evasion when the model becomes operational. Evasion attacks, on the other hand, leverage learned model behaviors by crafting adversarial examples that force misclassification through minor perturbations, which preserve semantic consistency while crossing decision boundaries. According to the taxonomy of machine learning security threats, attacks target either data integrity, exploit model vulnerabilities, or breach privacy by extracting sensitive information from trained models [6]. Those consist of privacy attacks towards machine learning systems using model inversion techniques, which reconstruct training data from model parameters, in addition to membership inference attacks that determine whether or not samples have participated in training a model. Minimizing these risks requires careful architecture design, continuous model validation, and integration of human expertise into automated detection workflows.

Threat Type	Attack Method	Target	Defence Strategy
Poisoning Attacks	Inject malicious samples into training data	Training data and learning process	Data validation; anomaly detection; robust algorithms
Evasion Attacks	Craft adversarial inputs with subtle perturbations	Deployed model inference	Adversarial training; input sanitisation; ensemble defences
Model	Reconstruct training data from	Model parameters	Differential privacy; output

Inversion	model parameters	and architecture	perturbation; limited access
Membership Inference	Determine if samples were in the training set	Training dataset and model responses	Privacy-preserving training; regularisation techniques

Table 3. Security Threats Targeting Machine Learning Systems [6]

Operational Architecture for AI-Network Security Integration

This requires data pipelines that gather telemetry from distributed sources, normalise heterogeneous formats, extract relevant features, and perform real-time inference, integrating the outputs into security operations workflows. Telemetry sources range from network flow records and domain name system query logs to proxy traffic data, virtual private network tunnel metadata, and endpoint sensor feeds. The architectural framework of AI-driven network security systems integrates several elements that have to work in harmony in order for operational effectiveness to be achieved. Data collection infrastructure has to accommodate a wide variety of telemetry sources operating at different sampling rates and producing heterogeneous data formats; this calls for normalization layers that transform raw inputs to standardized representations that can subsequently be used by downstream processing stages. Contemporary network environments, especially those where Internet of Medical Things devices and other specialized endpoints are integrated, produce heterogeneous patterns of traffic that require specialized processing capabilities. Research into intrusion detection in specialized network domains illustrates the idea that holistic datasets covering a wide range of attack scenarios, across several protocol layers, allow for substantial model training. Particular emphasis is placed on the capture of network-layer flows and application-layer interactions, which can reveal attack behaviors invisible at individual protocol strata [7]. Streaming platforms aggregate this information at low latency, allowing for near real-time processing through messaging queuing systems that decouple the data producers and consumers, providing some buffering capacity that accommodates temporary delays in processing without loss of data.

Feature engineering transforms raw packet-level data into statistical representations that include connection duration distributions, flow count aggregations, entropy measures of domain name strings, byte variance patterns, and temporal sequencing of protocol events. Relevant features can enable learning algorithms to identify discriminative patterns, whereas the presence of irrelevant and redundant features increases dimensionality without improving the detection capability. The construction of a dataset for training an intrusion detection model requires due consideration to attack diversity, protocol coverage, and realistic traffic generation methodologies that reflect operational network conditions rather than synthetic laboratory environments. Emphasis has been put on benchmark datasets targeting the capturing of complete network sessions across multiple layers of protocols to allow models to learn relationships between transport-layer flows and application-layer semantics that are characteristic of both normal operations and malicious activities [7]. Integration layers expose the detection outputs through application programming interfaces that connect with security orchestration platforms, thus driving automated response actions if the threat scores exceed configured thresholds.

Pivotal to continued effectiveness is the establishment of feedback mechanisms wherein security analysts label detection outcomes, enabling model retraining in a continuous cycle that adapts to emergent threat patterns while reducing false positives with each successive iteration. These machine learning systems have to be continuously validated and refined for implementation in security monitoring; continuous cycles that build from analyst feedback about detection accuracy are needed. Human-artificial intelligence collaboration frameworks recognize that while automated systems excel at processing large volumes of information to identify statistical patterns, human analysts provide necessary contextual interpretation, domain expertise, and reasoning capabilities beyond existing algorithmic capabilities. Augmented intelligence frameworks go beyond mere automation by instituting true partnership models whereby the expertise of humans and the capabilities of artificial intelligence interlink across the lifecycle of security operations. Research into human-artificial intelligence teaming within cybersecurity contexts demonstrates that such effective collaboration necessitates architectural frameworks that offer support for bidirectional information flow wherein the artificial intelligence systems do not only present findings about detections but solicit human guidance about uncertain classifications, contextual factors impacting the interpretation of threats, and strategic priorities informing response decisions [8]. This kind of collaboration paradigm demands interface designs that present artificial intelligence reasoning processes transparently so that analysts can make sense of the rationale underpinning detections to identify potential limitations or biases in models that need correction through retraining or adjustment in architecture [8].

Architecture Layer	Components	Processing Requirements	Integration Method
Telemetry Collection	Flow records, DNS logs, proxy data, VPN metadata, endpoint sensors	Handle diverse formats and high volumes; horizontal scaling	Streaming platforms, message queuing, and buffering systems
Feature Engineering	Statistical representations, duration metrics, entropy measures, and temporal patterns	Transform raw packets to discriminative features	Normalisation layers; automated deep learning extraction
Inference Engine	Anomaly scoring models; probability-based risk estimates	Millisecond-scale latency; multi-protocol processing	APIs for detection outputs; threshold-based automation
Human-AI Collaboration	Analyst feedback: active learning for uncertain cases	Bidirectional information exchange with experts	Transparent reasoning display; continuous retraining loops

Table 4. AI-Driven Security Operations Architecture Components [7, 8].

Conclusion

The integration of artificial intelligence capabilities into network defense infrastructures marks a fundamental evolution from reactive signature matching to proactive behavioral anomaly detection capable of identifying previously unknown threats. Traditional protection architectures designed for static, perimeter-centric environments cannot thoroughly address the modern threat landscape characterized by distributed cloud deployments, ubiquitous encryption, and adversaries leveraging artificial intelligence for offensive purposes. Machine learning methodologies enable the identification of malicious activities through statistical deviations from learned baseline patterns, instead of specific matching of predefined attack signatures, affording defensive capabilities against zero-day exploits and novel attack vectors for which prior threat intelligence does not exist. Successful operational deployment extends beyond algorithmic sophistication to include comprehensive data engineering pipelines capable of collecting and normalizing heterogeneous telemetry streams, feature extraction techniques that transform raw network observations into discriminative representations, and architectural designs that appropriately balance detection accuracy against latency constraints necessary for real-time traffic enforcement. Critical challenges consist of class imbalance, model interpretability, and adversarial robustness—continuous attention should be given to these factors through careful training data curation, explainable artificial intelligence techniques that render decision-making processes transparent to human analysts, and defensive mechanisms aimed at defending against poisoning attacks or evasive input crafting. Augmented intelligence frameworks recognize complementary strengths of automated systems and human expertise, thus establishing collaborative operational paradigms wherein algorithms process large-scale telemetry and domain experts provide contextual interpretation and strategic guidance. The trajectory towards adaptive, self-learning protection systems embedded within cloud-native security architectures reflects a growing recognition that effectively defending increasingly complex network environments against sophisticated adversaries requires equal sophistication in defensive capabilities, shifting beyond static rules towards dynamic systems that continuously learn, adapt, and evolve through operational experience.

References

[1] Siffat Ullah Khan et al., "Challenges and Their Practices in Adoption of Hybrid Cloud Computing: An Analytical Hierarchy Approach," Wiley, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2021/1024139

[2] Bhagwant Singh and Sikander Singh Cheema, "Emerging Trends in AI-Powered Malware Detection: A Review of Real-Time and Adversarially Resilient Techniques," ResearchGate, 2024. [Online]. Available: https://www.researchgate.net/profile/Bhagwant-Singh-4/publication/388405244

[3] Ansam Khraisat et al., "Survey of intrusion detection systems: techniques, datasets and challenges," Cybersecurity, 2019. [Online]. Available: https://link.springer.com/content/pdf/10.1186/s42400-019-0038-7.pdf

[4] Tal Shapira and Yuval Shavitt, "FlowPic: Encrypted Internet Traffic Classification is as Easy as Image Recognition," IEEE INFOCOM WKSHPS, 2019, pp. 680-687. [Online]. Available: https://www.researchgate.net/profile/Yuval-Shavitt/publication/336010067_FlowPic_Encrypted_Internet_Traffic_Classification_is_as_Easy_as_Image_Recognition/links/61c2fd5b52bd3c7e0583c60a/FlowPic-Encrypted-Internet-Traffic-Classification-is-as-Easy-as-Image-Recognition.pdf

[5] Vanlalruata Hnamte and Jamal Hussain, "DCNNBiLSTM: An Efficient Hybrid Deep Learning-Based Intrusion Detection System," ScienceDirect, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2772503023000130

[6] QIANG LIU et al., "A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View," IEEE Access, 2018. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8290925

[7] JOSÉ AREIA, "IoMT-TrafficData: Dataset and Tools for Benchmarking Intrusion Detection in Internet of Medical Things," IEEE Access, 2024. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10620207

[8] Masike Malatji, "Augmented Intelligence Framework for Human–Artificial Intelligence Teaming in Cybersecurity," Human-Centric Intelligent Systems, 2025. [Online]. Available: https://link.springer.com/content/pdf/10.1007/s44230-025-00103-8.pdf