International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

A Hybrid Model for Abnormal Network Traffic Identification and Threat Detection

Onuma Divine Anele[1], PROF. D. Matthias[2], DR. O. E. Taylor[3], PROF. N. D. Nwiabu[4]

[1,2,3,4]*Department of Computer Science, Rivers State University, Nkpolu-Oroworukwo, Port Harcourt, Rivers State, Nigeria.*

***Abstract*— The increasing sophistication of cyberattacks and the widespread adoption of encrypted network communications have exposed significant limitations in conventional intrusion detection systems, particularly in detecting zero-day attacks while preserving data privacy. Existing approaches often suffer from reduced detection accuracy, high false positive rates, increased detection latency, and inadequate privacy protection when analyzing encrypted network traffic. This study proposes a privacy-preserving hybrid framework for abnormal network traffic identification by integrating Long Short-Term Memory (LSTM), Random Forest (RF), K-Means clustering, and Cheon–Kim–Kim–Song (CKKS) homomorphic encryption. The framework was designed to leverage LSTM for temporal traffic modeling, Random Forest for supervised classification, K-Means clustering for unsupervised anomaly discovery, and CKKS encryption for secure computation on encrypted network data. The proposed model was implemented in Python and evaluated using the CICIDS2017 and UNSW-NB15 benchmark datasets. Performance was assessed using accuracy, precision, recall, F1-score, AUC-ROC, detection rate, false positive rate, detection latency, throughput, scalability, adversarial robustness, explainability, privacy preservation, computational overhead, encryption overhead, and zero-day attack detection. Experimental results demonstrated that the proposed framework consistently outperformed the baseline model of Sattar [133]. On the CICIDS2017 dataset, the model achieved 99.31% accuracy, 99.12% precision, 99.05% recall, 99.08% F1-score, 99.05% detection rate, and 92.4% zero-day detection, while reducing the false positive rate to 0.74% and detection latency to 18.7 ms. Similar improvements were obtained on the UNSW-NB15 dataset, with 98.05% accuracy, 97.72% precision, 97.55% recall, 97.63% F1-score, and 89.1% zero-day detection, alongside lower computational and encryption overheads and enhanced scalability, explainability, adversarial robustness, and privacy preservation. The study concludes that integrating deep learning, machine learning, unsupervised clustering, and homomorphic encryption provides a scalable, accurate, and privacy-preserving solution for real-time abnormal network traffic identification. The proposed framework offers a practical approach for deployment in cloud computing, IoT, healthcare, financial systems, telecommunications, and other security-critical network environments.**

***Keywords*— Abnormal Network Traffic Identification, Intrusion Detection System, LSTM, Random Forest, K-Means Clustering, CKKS Homomorphic Encryption, Privacy-Preserving Machine Learning, Zero-Day Attack Detection, Network Security.**

I. INTRODUCTION

The increasing adoption of cloud computing, the Internet of Things (IoT), distributed systems, and encrypted communication has significantly expanded the complexity of modern computer networks and the sophistication of cyber threats. Network traffic, comprising packets exchanged among interconnected devices, provides valuable behavioral information that can be analyzed to distinguish legitimate communication from malicious activities. Abnormal network traffic, characterized by irregular packet flows, unusual protocol usage, anomalous communication patterns, and excessive bandwidth consumption, often serves as an early indicator of cyberattacks such as malware infections, botnet communications, distributed denial-of-service (DDoS) attacks, brute-force attempts, and data exfiltration. Consequently, abnormal network traffic identification has become a critical component of modern intrusion detection systems (IDSs).

Conventional signature-based intrusion detection techniques have proven inadequate against rapidly evolving threats, particularly zero-day attacks and encrypted network traffic. Recent machine learning and deep learning approaches, including Long Short-Term Memory (LSTM), Random Forest, K-Means clustering, and self-supervised learning, have substantially improved anomaly detection by exploiting sequential, statistical, and behavioral traffic features. However, existing methods continue to face several challenges, including high false-positive rates, limited adversarial robustness, inadequate detection of previously unseen attacks, scalability constraints in high-volume heterogeneous networks, and insufficient privacy protection during traffic analysis.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Although the encrypted traffic self-supervised learning (ET-SSL) framework proposed by **Sattar [133]** achieved promising detection accuracy, it does not integrate formal privacy-preserving cryptographic mechanisms and relies primarily on a single learning paradigm, limiting its ability to jointly exploit temporal dependencies, supervised classification, and unsupervised anomaly discovery.

These limitations highlight the need for a scalable, intelligent, and privacy-preserving anomaly detection framework capable of accurately identifying both known and emerging cyber threats while protecting sensitive network information. This study addresses these challenges by developing a hybrid anomaly detection model that integrates LSTM, Random Forest, K-Means clustering, and Cheon–Kim–Kim–Song (CKKS) homomorphic encryption. The proposed framework combines sequential learning, supervised classification, unsupervised clustering, and privacy-preserving computation to improve detection accuracy, enhance adversarial robustness, reduce false positives, and enable secure analysis of encrypted network traffic.

Accordingly, the aim of this study is to develop an efficient hybrid model for abnormal network traffic identification and threat detection that integrates supervised and unsupervised artificial intelligence techniques with homomorphic encryption for improved accuracy, scalability, and privacy preservation. Specifically, the study seeks to: (i) design and implement a scalable hybrid anomaly detection framework integrating LSTM, Random Forest, and K-Means clustering; (ii) enhance adversarial robustness by combining complementary learning paradigms; (iii) improve the detection of evolving and zero-day attacks in encrypted network environments while minimizing false positives; (iv) integrate CKKS homomorphic encryption to enable secure computation on encrypted traffic; and (v) evaluate the proposed framework using standard performance metrics, including accuracy, precision, recall, F1-score, false-positive rate, detection latency, and throughput under real-time and adversarial conditions.

II. Literature Review

2.1 Conceptual Framework of Network Traffic

Network traffic analysis provides the foundation for monitoring, optimizing, and securing modern communication networks by examining the behavior of packets, flows, and events transmitted across interconnected systems.

Effective traffic analysis relies on understanding the sources of network data, traffic characteristics, and environmental factors influencing communication behavior. These components collectively support anomaly detection, network performance optimization, and cybersecurity operations.

Network traffic information is obtained from several complementary data sources, including packet-level data, flow-based records, system log files, network management systems, and real-time traffic streams. Packet-level data provide detailed information such as source and destination addresses, protocols, ports, and payloads, enabling fine-grained analysis of malware communication, botnets, and data exfiltration **[167, 10]**. However, their large volume and privacy implications increase storage and computational requirements **[80, 44]**. Flow-based data summarize communication sessions using protocols such as NetFlow and sFlow, supporting scalable traffic monitoring and attack detection, although aggregation may obscure fine-grained attack characteristics **[102, 74, 34]**. Log files generated by routers, firewalls, and intrusion detection systems provide contextual information for forensic analysis but require extensive preprocessing because of their volume and complexity **[148, 167, 41]**. Similarly, network management systems provide operational metrics such as throughput, device availability, and error rates, while real-time data streams enable continuous monitoring and rapid response to emerging threats despite their significant computational demands **[8, 33, 142, 167]**.

Traffic characteristics constitute another essential component of network traffic analysis. Important features include traffic volume, flow duration, inter-arrival time, packet size distribution, temporal behavior, protocol usage, burstiness, and source–destination communication patterns. Traffic volume is widely used to identify congestion and denial-of-service attacks **[147, 57, 74]**. Temporal characteristics such as flow duration and packet inter-arrival time reveal slow-rate attacks and reconnaissance activities **[56]**, whereas packet size distributions and temporal periodicity assist in detecting fragmentation attacks, data exfiltration, and abnormal communication behaviors **[153, 83, 148]**. Likewise, protocol usage patterns, traffic burstiness, self-similarity, and source–destination attributes provide valuable indicators of malicious activities including botnet communications and unauthorized network access **[56, 120, 71]**.

The effectiveness of network traffic analysis is influenced by several operational factors, including network topology, traffic volume, protocol diversity, encryption, user behavior, and hardware capabilities.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Large-scale heterogeneous networks generate diverse traffic patterns that complicate monitoring and anomaly detection, while widespread encryption limits payload inspection and shifts detection toward metadata- and flow-based analysis **[33]**. Consequently, modern traffic analysis increasingly depends on intelligent, adaptive, and scalable analytical techniques capable of maintaining high detection accuracy in dynamic network environments.

2.2 Abnormal Network Traffic Identification

Abnormal network traffic identification is fundamental to modern cybersecurity because it enables the early detection of malicious activities that threaten the confidentiality, integrity, and availability of network resources. Abnormal traffic refers to any deviation from established communication patterns, including unusual traffic volumes, unauthorized access attempts, abnormal protocol usage, data exfiltration, malware communication, and Denial-of-Service (DoS) attacks **[142, 29]**. As enterprise, cloud, and IoT networks continue to expand, timely identification of these anomalies has become increasingly important for preventing cyberattacks, minimizing operational disruption, and maintaining regulatory compliance.

The significance of abnormal traffic detection extends beyond attack identification. Early detection enables organizations to mitigate Distributed Denial-of-Service (DDoS) attacks, malware propagation, insider threats, and unauthorized data transfers before they escalate into large-scale security incidents **[167, 55]**. It also protects sensitive information from data breaches, maintains network availability and Quality of Service (QoS), supports compliance with security regulations such as GDPR, HIPAA, and PCI DSS, facilitates rapid incident response, and improves the overall security posture by reducing false alarms and strengthening proactive defense mechanisms **[29, 96, 142]**. Despite these advantages, abnormal traffic detection remains challenging because of increasing traffic volumes, encrypted communications, evolving attack techniques, and the need for accurate real-time analysis.

Several approaches have been developed for abnormal network traffic identification, each offering distinct advantages and limitations. Statistical methods establish baseline traffic profiles and identify deviations using predefined thresholds; however, they often struggle with complex and dynamic network environments **[96]**. Signature-based or misuse detection identifies attacks by matching network traffic against known attack signatures and is highly effective for detecting previously documented threats.

Nevertheless, its dependence on continually updated signature databases limits its ability to detect zero-day attacks, polymorphic malware, and other previously unseen threats **[142, 167, 29]**. Rule-based and heuristic approaches similarly provide efficient detection for predefined attack patterns but frequently suffer from high false-positive rates in heterogeneous environments **[96]**.

Anomaly detection addresses many of these limitations by learning normal network behavior and identifying deviations without relying on predefined attack signatures. This capability makes anomaly detection particularly effective for identifying novel attacks, advanced persistent threats, and zero-day exploits **[96, 167]**. Statistical anomaly detection techniques model normal traffic distributions using approaches such as Gaussian Mixture Models, while machine learning methods employ supervised algorithms including Random Forest, Support Vector Machines (SVM), and Neural Networks, as well as unsupervised algorithms such as K-Means clustering, Isolation Forest, and Autoencoders **[29, 55]**. More recently, deep learning architectures such as Long Short-Term Memory (LSTM) networks have demonstrated superior performance by learning long-term temporal dependencies within sequential network traffic, enabling the detection of stealthy and multi-stage attacks **[167]**. Behavioral analysis further complements anomaly detection by monitoring deviations in user and device activities, making it particularly useful for identifying insider threats **[142]**.

Although these techniques have significantly improved intrusion detection, each possesses inherent limitations. Signature-based systems cannot detect unknown attacks, statistical approaches often produce excessive false positives, and standalone machine learning models generally fail to simultaneously capture temporal dependencies, complex feature interactions, and evolving attack behaviors **[130, 121, 163]**. Likewise, K-Means clustering struggles with high-dimensional noisy data and requires prior specification of the number of clusters, while traditional supervised models exhibit limited adaptability to changing attack patterns **[148]**. These shortcomings reduce their effectiveness in large-scale, real-time environments characterized by encrypted traffic and continuously evolving cyber threats.

To overcome these limitations, recent studies advocate hybrid detection architectures that combine complementary learning paradigms. Integrating K-Means clustering, Random Forest classification, and LSTM networks enables simultaneous exploitation of unsupervised pattern discovery, supervised classification, and temporal sequence learning.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

K-Means facilitates the discovery of previously unknown attack patterns, Random Forest provides robust and interpretable classification, while LSTM captures temporal relationships necessary for detecting stealthy, coordinated, and multi-stage intrusions **[24, 61, 148]**. Ensemble decision fusion further improves detection confidence by combining outputs from multiple models, thereby increasing detection accuracy while reducing false-positive rates.

Privacy preservation has also become a major consideration in intelligent network traffic analysis because network data frequently contain sensitive organizational and personal information. Contemporary detection frameworks increasingly integrate privacy-preserving mechanisms including data anonymization, federated learning, differential privacy, homomorphic encryption, and Secure Multi-Party Computation (SMPC) to enable collaborative learning and secure analysis without exposing confidential data **[46, 103, 43, 2, 113]**. These approaches enhance legal compliance while supporting distributed deployment across cloud, enterprise, healthcare, financial, and critical infrastructure environments.

The effectiveness of hybrid anomaly detection systems is commonly evaluated using benchmark datasets such as NSL-KDD, CICIDS2017, and UNSW-NB15, with performance measured using accuracy, precision, recall, F1-score, false-positive rate, and detection latency. Comparative evaluations consistently demonstrate that hybrid frameworks outperform standalone machine learning models by improving zero-day attack detection, reducing false positives, and enhancing temporal sensitivity. Furthermore, scalable implementations based on distributed processing frameworks and containerized deployment architectures enable efficient operation across enterprise, cloud, and IoT environments while maintaining high detection accuracy and low latency **[159, 138, 118, 136]**.

2.3 Hybrid Models for Abnormal Network Traffic Identification

Hybrid machine learning models have emerged as an effective approach for abnormal network traffic identification by combining the complementary strengths of unsupervised learning, supervised learning, and deep learning techniques. Unlike standalone detection models, hybrid architectures simultaneously exploit pattern discovery, classification accuracy, and temporal sequence learning, thereby improving the detection of both known and previously unseen cyber threats.

This integration addresses key limitations of conventional intrusion detection systems, including high false-positive rates, poor adaptability to evolving attacks, and limited capability for analyzing complex traffic patterns **[28, 74]**.

K-Means clustering serves as the unsupervised component of many hybrid frameworks by partitioning network traffic into groups based on feature similarity without requiring labeled data. This capability enables the discovery of hidden structures and emerging attack patterns within large-scale traffic datasets, making it particularly valuable for identifying zero-day attacks and reducing noise before supervised classification. However, K-Means alone cannot accurately distinguish malicious from legitimate traffic because it lacks class labels and is sensitive to cluster initialization and high-dimensional data **[68, 28]**.

Random Forest provides the supervised learning component by constructing multiple decision trees and aggregating their predictions to produce robust classifications. The algorithm is widely adopted in network intrusion detection because of its ability to process high-dimensional traffic features, resist overfitting, rank feature importance, and maintain high classification accuracy across diverse attack categories **[24, 74]**. Nevertheless, Random Forest analyzes traffic instances independently and therefore cannot effectively capture temporal dependencies or sequential attack behaviors commonly observed in sophisticated cyberattacks.

Long Short-Term Memory (LSTM) networks complement Random Forest by modeling the temporal characteristics of network traffic. As an advanced recurrent neural network architecture, LSTM retains long-term dependencies through gated memory mechanisms, enabling it to recognize evolving attack sequences, stealthy intrusions, and low-rate attacks that static classifiers may overlook **[61]**. Because network traffic naturally exhibits sequential behavior, LSTM has become an effective tool for detecting time-dependent anomalies and multi-stage cyberattacks **[74]**.

The integration of K-Means, Random Forest, and LSTM into a unified hybrid framework leverages the strengths of each technique while mitigating their individual limitations. K-Means performs initial clustering and anomaly discovery, Random Forest accurately classifies labeled traffic, and LSTM models temporal evolution within traffic flows. This multi-layered architecture enhances detection accuracy, improves zero-day attack identification, reduces false-positive rates, and strengthens model generalization across heterogeneous network environments **[17, 74]**.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Furthermore, the modular design allows each component to be independently optimized or retrained without affecting the overall framework, thereby improving scalability, maintainability, and deployment flexibility in real-time intrusion detection systems.

2.4 Related Works

Machine learning has become a dominant approach for abnormal network traffic detection because of its ability to learn complex traffic patterns and improve intrusion detection accuracy. Existing studies broadly classify detection methods into supervised, unsupervised, semi-supervised, and hybrid learning approaches **[140]**. Supervised learning models, including Neural Networks **[155]**, Support Vector Machines **[107]**, Decision Trees **[146]**, Naïve Bayes **[7]**, and ensemble techniques such as Random Forest **[119]**, have demonstrated high classification accuracy when sufficient labeled data are available. However, their performance is constrained by the availability of annotated datasets and limited adaptability to evolving attack patterns. Conversely, unsupervised approaches, particularly clustering techniques, identify latent traffic structures without labeled data by exploiting similarity measures, making them suitable for detecting unknown attacks, although they often suffer from lower classification precision **[123]**.

Semi-supervised learning combines labeled and unlabeled data to improve generalization and detection performance. **Chen [30]** demonstrated that semi-supervised techniques based on spectral graph transducers and Gaussian fields significantly improve classification accuracy compared with purely supervised approaches. Furthermore, feature selection and dimensionality reduction have become essential preprocessing steps for reducing redundancy and computational complexity. For example, **Li [93]** proposed an improved Krill Swarm optimization algorithm for feature selection, while **Zhang [160]** employed an autoencoder to learn compressed traffic representations before intrusion classification, thereby enhancing detection efficiency.

Recent studies have focused on improving detection accuracy using advanced machine learning techniques. **Bian [18]** combined statistical analysis with Isolation Forest to detect anomalous traffic, achieving precision, recall, and F1-scores above 0.98. Nevertheless, the approach depends on fixed statistical thresholds, ignores temporal dependencies in traffic behavior, and was validated on a relatively small dataset, limiting its scalability to real-world environments.

Similarly, **Chen [31]** enhanced Decision Tree-based intrusion detection through data normalization and Principal Component Analysis (PCA), reporting improved detection accuracy and processing speed on the NSL-KDD dataset. However, reliance on traditional Decision Trees and PCA restricts the model's ability to capture sequential traffic characteristics, adapt to sophisticated attack behaviors, and generalize to modern encrypted or high-volume network environments. Support Vector Machine (SVM)-based intrusion detection has also demonstrated competitive performance. **Kong [77]** developed an Abnormal Traffic Identification System (ATIS) using SVM and highlighted the importance of data normalization and parameter optimization in improving classification performance. Despite its effectiveness, SVM requires extensive labeled datasets, exhibits limited scalability for multi-class classification, and lacks the capability to model temporal dependencies and continuously evolving attack behaviors.

To better exploit temporal information, **Li [91]** introduced the Random Shapelet Forest for Intrusion Detection (RSFID), integrating Fast Correlation-Based Filter (FCBF) feature selection with Symbolic Aggregate approXimation (SAX) to improve time-series intrusion detection. The proposed method effectively captured temporal traffic characteristics while maintaining model interpretability. However, its dependence on static shapelet sampling and benchmark datasets limits its applicability to high-dimensional, real-time network traffic and does not address privacy-preserving analysis.

Overall, the literature demonstrates substantial progress in machine learning-based intrusion detection. Nevertheless, existing approaches predominantly rely on a single learning paradigm and therefore struggle to simultaneously exploit spatial, temporal, and clustering characteristics of network traffic. They also exhibit limitations in detecting zero-day attacks, reducing false positives, scaling to high-speed heterogeneous networks, and preserving data privacy during analysis. These shortcomings motivate the development of a hybrid framework that integrates K-Means clustering for anomaly discovery, Random Forest for robust supervised classification, Long Short-Term Memory (LSTM) for temporal sequence learning, and CKKS homomorphic encryption for privacy-preserving anomaly detection in encrypted network environments.

Adhikari [4] presented a comprehensive survey of anomaly detection techniques for the Internet of Things (IoT), emphasizing the limitations imposed by resource-constrained IoT devices on conventional detection approaches.

The survey reviews statistical, machine learning, and hybrid methods across applications including network security, fraud detection, and industrial automation.

The authors identified challenges such as the absence of cross-layer detection mechanisms, limited standardized benchmark datasets, and insufficient real-time detection techniques for constrained environments. They recommended lightweight, scalable, and context-aware anomaly detection models to improve IoT security, highlighting the need for adaptive and efficient detection frameworks **[4]**.

Almuhanna and Dardouri [9] proposed a hybrid Network Intrusion Detection System (NIDS) that combines XGBoost, Random Forest, Graph Neural Networks (GNN), Long Short-Term Memory (LSTM), and Autoencoders within a weighted soft-voting ensemble. Using more than 5.6 million network traffic records, the framework incorporates feature engineering and SMOTE to address class imbalance, achieving high detection accuracy and strong generalization across multiple intrusion scenarios. Although the ensemble significantly improves anomaly detection compared with traditional signature-based methods, the study does not comprehensively evaluate real-time deployment, model interpretability, or performance across heterogeneous operational environments **[9]**.

Chentoufi [35] developed a machine learning-based intrusion detection framework using the UNSW-NB15 dataset. The study evaluated Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, and XGBoost after preprocessing, SMOTE oversampling, and Principal Component Analysis (PCA). Random Forest achieved the best performance with an F1-score of 97.80% and a false alarm rate of 1.37%, demonstrating the effectiveness of combining feature engineering, dimensionality reduction, and class balancing for intrusion detection. However, the framework was validated on a single dataset, and the authors recommended evaluation on additional benchmark datasets and the integration of hybrid or deep feature learning techniques to improve robustness and generalization **[35]**.

Koukoulis [78] introduced a self-supervised contrastive learning framework based on transformer encoders for intrusion detection without requiring labeled data. The proposed approach learns packet-level representations directly from raw traffic sequences using packet augmentation techniques instead of handcrafted NetFlow features. Experimental results demonstrated improvements of up to 3% in intra-dataset Area Under the Curve (AUC) and 20% in inter-dataset AUC over existing self-supervised approaches while providing a strong foundation for supervised intrusion detection with limited labeled samples. Nevertheless, the framework requires further investigation regarding deployment efficiency, computational cost, interpretability, and performance in heterogeneous network environments **[78]**.

Guerra [59] proposed a self-supervised graph representation learning approach that models network traffic as graphs to capture structural relationships among hosts and communication flows. By generating graph embeddings that preserve topological information, the proposed framework improves anomaly separability and achieves competitive intrusion detection performance compared with conventional self-supervised methods. Despite these promising results, challenges remain regarding computational scalability, real-time inference, and validation across large-scale heterogeneous networks and diverse attack categories **[59]**.

Yang [156] developed a Large Language Model (LLM)-based framework for cloud network traffic monitoring and anomaly detection. The framework employs transformer-based language models to process raw traffic sequences and integrates transfer learning to improve adaptation to unseen cloud environments and adversarial conditions. Experimental results indicated superior accuracy, reduced false positives, and improved detection of zero-day attacks compared with conventional machine learning approaches. However, questions remain concerning scalability, explainability, and robustness across heterogeneous cloud infrastructures, highlighting the need for further optimization of LLM-based cybersecurity systems **[156]**.

Nong [114] investigated statistical deviation-based anomaly detection for highly imbalanced network traffic data. The proposed framework detects anomalies by identifying significant deviations from established traffic baselines, enabling reliable identification of rare events such as Distributed Denial-of-Service (DDoS) attacks and flash crowds without relying heavily on labeled datasets. While the approach enhances monitoring robustness under class imbalance, further research is required to integrate statistical deviation techniques with machine learning and deep learning models, evaluate performance under varying imbalance ratios, and develop adaptive threshold mechanisms for dynamic network environments **[114]**.

Overall, recent studies demonstrate substantial progress in applying hybrid learning, self-supervised learning, graph neural networks, transformer architectures, and statistical techniques to abnormal network traffic identification. Nevertheless, important research gaps remain, including limited integration of sequential learning, supervised classification, unsupervised anomaly discovery, and privacy-preserving computation within a unified framework.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

In addition, existing studies provide limited evaluation of scalability, real-time deployment, adversarial robustness, interpretability, and secure processing of encrypted network traffic. These limitations motivate the development of a scalable hybrid framework that integrates Long Short-Term Memory (LSTM), Random Forest, K-Means clustering, and CKKS homomorphic encryption to improve detection accuracy, reduce false positives, enhance zero-day attack detection, and preserve data privacy in modern network environments.

Prabowo [124] evaluated the robustness of anomaly-based Network Intrusion Detection Systems (NIDSs) under imperfect training conditions by introducing controlled malicious traffic into benign training datasets. Five anomaly detection models—Local Outlier Factor (LOF), Isolation Forest, One-Class Support Vector Machine (OC-SVM), Autoencoders, and Long Short-Term Memory (LSTM)—were assessed using HTTP, FTP, and SMTP traffic. The results showed that Autoencoders based on byte-frequency features achieved the highest robustness, maintaining superior F2 scores despite increasing contamination, while classical algorithms were more susceptible to noisy datasets and LSTM demonstrated intermediate performance. Although the study provides valuable guidance for selecting anomaly detection models in environments with contaminated datasets, it does not comprehensively evaluate computational efficiency, inference latency, memory consumption, or scalability for large-scale real-time deployment, leaving important practical implementation issues unresolved **[124]**.

Prasad [125] investigated the effectiveness of conventional machine learning algorithms for network anomaly detection by formulating intrusion detection as a classification problem. The study compared Support Vector Machines (SVM), Decision Trees, k-Nearest Neighbors (k-NN), Naïve Bayes, and Random Forest using feature normalization and selection techniques prior to model training. Random Forest consistently achieved superior performance in terms of accuracy, precision, recall, and F1-score, demonstrating the continued relevance of traditional machine learning methods for intrusion detection. However, the proposed framework relies heavily on labeled datasets, is sensitive to class imbalance, and exhibits limited generalizability to encrypted and heterogeneous network traffic, indicating the need for adaptive hybrid and deep learning architectures **[125]**.

Palaniappan [117] examined machine learning approaches for insider threat detection by comparing Support Vector Machines, Random Forest, Gradient Boosting, k-Nearest Neighbors, and Naïve Bayes using network flow datasets.

Ensemble methods, particularly Random Forest and Gradient Boosting, achieved superior precision, recall, F1-score, and Area Under the Curve (AUC), demonstrating their effectiveness in identifying subtle insider threats embedded within legitimate user activities. Nevertheless, the study was limited to curated datasets and did not investigate large-scale operational environments, temporal behavior modeling, concept drift, or hybrid deep learning approaches capable of capturing evolving insider attack patterns **[117]**.

Ibrahim [67] conducted a systematic review of machine learning techniques for network anomaly detection, covering supervised, unsupervised, and semi-supervised approaches including Support Vector Machines, Random Forests, clustering methods, k-Nearest Neighbors, and autoencoder-based models. The review concluded that ensemble and hybrid learning models generally outperform standalone classifiers in terms of detection accuracy and robustness while highlighting persistent challenges such as class imbalance, noisy data, high-dimensional feature spaces, and elevated false alarm rates in unsupervised models. The authors further identified the absence of standardized benchmarking, limited real-time deployment evaluations, and insufficient adaptive learning mechanisms as key research gaps requiring future investigation **[67]**.

Oziegbe [115] explored anomaly-based intrusion detection for Vehicular Ad Hoc Networks (VANETs) through hybrid architectures that integrate deep learning models with conventional machine learning algorithms. The study demonstrated that combining deep neural networks, convolutional neural networks, and recurrent architectures with classical classifiers significantly improves the detection of zero-day attacks and complex intrusion behaviors within highly dynamic vehicular environments. Despite improved detection accuracy, the framework remains constrained by high computational complexity, limited interpretability, and scalability challenges associated with resource-constrained automotive systems. The study therefore recommends lightweight hybrid learning architectures and extensive real-world validation using diverse vehicular traffic datasets **[115]**.

Early Anomaly Detection in Network Traffic Using Deep Methods [45] demonstrated the effectiveness of deep learning techniques, including Autoencoders, Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and Convolutional Neural Networks (CNNs), for early anomaly detection in network traffic logs.

By automatically learning hierarchical representations from raw traffic data, these models improved detection sensitivity and reduced false alarm rates compared with traditional statistical and shallow machine learning methods. The study also highlighted the advantages of sequence embedding, time-series modeling, and unsupervised deep clustering in environments with limited labeled data. However, challenges remain regarding computational overhead, inference latency, cross-dataset generalization, and comprehensive comparisons with recent self-supervised and contrastive learning frameworks for practical real-time deployment **[45]**.

Collectively, these studies demonstrate that ensemble learning, deep learning, and hybrid detection frameworks consistently outperform conventional standalone approaches in abnormal network traffic identification. Nevertheless, common limitations persist, including inadequate scalability evaluation, limited privacy-preserving mechanisms, high computational costs, sensitivity to evolving attack patterns, dependence on labeled datasets, and insufficient validation under heterogeneous real-world network environments. These research gaps motivate the development of a scalable hybrid framework that integrates K-Means clustering, Random Forest, Long Short-Term Memory (LSTM), and CKKS homomorphic encryption to enhance detection accuracy, adversarial robustness, privacy preservation, and real-time operational performance.

siForest [137] proposed a set-aware Isolation Forest (siForest) that preserves intrinsic relationships within network traffic by terminating tree partitioning when all instances in a node share a common IP address and computing anomaly scores at the set level rather than the individual instance level. This approach enhances the detection of both sporadic and persistent anomalies in multidimensional traffic and demonstrated superior performance over conventional Isolation Forest methods on synthetic network scan datasets. However, the evaluation was limited to synthetic data, leaving unresolved issues regarding scalability, generalization to heterogeneous network environments, and effectiveness in encrypted or highly dynamic traffic scenarios **[137]**.

Miguel-Diez [101] conducted a systematic literature review on unsupervised learning algorithms for flow-based anomalous traffic detection. The review highlights the effectiveness of clustering, density-based methods, principal component analysis, autoencoders, and self-organizing maps in identifying anomalies without requiring labeled datasets. The authors emphasize that flow-level analysis is particularly valuable where labeled intrusion data are scarce.

Nevertheless, they identified persistent challenges, including high false-positive rates, sensitivity to feature selection and scaling, inconsistent benchmarking across datasets, and the absence of standardized evaluation frameworks. They therefore recommend hybrid approaches that combine unsupervised learning with adaptive thresholding and semi-supervised refinement to improve detection robustness, interpretability, and real-world applicability **[101]**.

Liu [95] examined machine learning-based network anomaly detection for strengthening cybersecurity defense mechanisms. The study synthesized supervised, unsupervised, and hybrid learning techniques, including clustering algorithms, statistical models, ensemble classifiers, and deep neural networks, demonstrating their effectiveness in detecting deviations from normal traffic behavior beyond the capabilities of traditional signature-based systems. Despite these advances, the authors reported challenges associated with data imbalance, high false-positive rates, limited model interpretability, and the absence of standardized benchmarks across heterogeneous network environments. They further emphasized the need for scalable, real-time, and explainable anomaly detection frameworks integrated with broader cybersecurity mechanisms such as encryption and incident response **[95]**.

Qu [126] proposed a network anomaly detection model based on event key time subgraphs. Instead of relying solely on conventional traffic features, the model represents network traffic as interconnected temporal event graphs, enabling the identification of coordinated and stealthy attacks through deviations in temporal connectivity patterns. Experimental evaluation on benchmark datasets demonstrated improved sensitivity and robustness compared with conventional approaches, particularly for anomalies exhibiting temporal dependencies. However, the computational overhead associated with subgraph construction and analysis limits scalability for high-speed networks, highlighting the need for optimization techniques and adaptive learning strategies to support real-time deployment in large-scale and heterogeneous environments **[126]**.

Santoso [132] investigated the application of the Naïve Bayes classifier for network anomaly detection using probabilistic modeling and k-fold cross-validation to distinguish normal from abnormal network traffic. The model was evaluated using accuracy, precision, recall, and F1-score, demonstrating that Naïve Bayes achieves competitive detection performance when trained on well-structured and discriminative features.

The authors highlighted its computational efficiency and interpretability, making it suitable for real-time network monitoring. However, its strong feature independence assumption and reduced performance on highly imbalanced datasets limit its applicability in complex network environments. The study recommended improved feature engineering and the development of hybrid or ensemble models to better capture feature interdependencies and enhance detection robustness **[132]**.

Wang [154] proposed the Residual Multi-scale Convolutional Learning with Attention (RMCLA) model for Internet of Things (IoT) intrusion detection. The framework transforms raw network traffic into structured feature images for convolutional neural network (CNN) processing while incorporating attention mechanisms and residual connections to improve feature extraction and classification. Evaluation on the CIC-IDS2017 and UNSW-NB15 datasets demonstrated superior detection accuracy and robustness compared with existing deep learning models. Nevertheless, challenges remain regarding computational complexity, scalability for real-time IoT deployment, and the limited interpretability of attention-based deep learning models, highlighting the need for lightweight and explainable architectures **[154]**.

Hozouri [63] presented a comprehensive survey of anomaly-based intrusion detection systems, reviewing statistical, machine learning, deep learning, and hybrid detection approaches across enterprise, cloud, and IoT environments. The survey emphasized the advantages of anomaly detection over signature-based methods in identifying unknown attacks and highlighted the increasing adoption of deep neural networks, ensemble learning, and unsupervised representations for improving detection performance. Despite these advances, significant challenges remain, including class imbalance, dataset scarcity, high false-positive rates, limited scalability, and insufficient model explainability. The authors further identified the lack of standardized benchmark datasets, limited comparative evaluations across heterogeneous traffic environments, and inadequate development of adaptive intrusion detection frameworks capable of responding to evolving cyber threats **[63]**.

Overall, these studies demonstrate substantial progress in applying probabilistic learning, deep learning, explainable artificial intelligence, and hybrid machine learning techniques for abnormal network traffic identification. Nevertheless, common limitations persist, including inadequate real-time scalability, computational complexity, limited interpretability, vulnerability to imbalanced and evolving datasets, and insufficient privacy-preserving capabilities.

These gaps justify the development of a scalable, interpretable, and privacy-preserving hybrid anomaly detection framework that integrates complementary learning paradigms to improve detection accuracy, reduce false positives, and enhance robustness against emerging cyber threats.

Lu [98] proposed a hybrid deep learning framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks for abnormal network traffic detection. The CNN component extracts hierarchical spatial features from network traffic, while the LSTM models temporal dependencies across sequential traffic flows, enabling the detection of both instantaneous and evolving anomalies. Experimental results demonstrated improved detection accuracy and lower false-positive rates compared with standalone models, confirming the effectiveness of joint spatial-temporal learning. However, the framework was evaluated primarily on offline benchmark datasets, leaving issues of real-time scalability, concept drift adaptation, and computational efficiency in high-throughput environments largely unexplored **[98]**.

Roy [131] presented a comprehensive survey of deep learning techniques for network traffic anomaly detection, reviewing architectures such as CNNs, Recurrent Neural Networks (RNNs), autoencoders, and hybrid deep learning models. The survey highlighted that deep learning significantly outperforms traditional statistical and signature-based approaches by effectively learning complex spatial and temporal traffic patterns. It further identified hybrid and ensemble models as achieving superior detection performance but emphasized persistent challenges related to data imbalance, computational overhead, model interpretability, and deployment in resource-constrained or high-speed environments **[131]**.

Jakkani [69] developed a real-time machine learning framework for network traffic anomaly detection using supervised algorithms including Random Forest and Support Vector Machines (SVM) on the UNSW-NB15 dataset. The study incorporated data preprocessing, feature selection, model training, and edge-computing architectures to improve scalability and reduce detection latency. Although the framework achieved high accuracy, precision, and recall, challenges remained in robust feature extraction, adversarial resilience, and adaptive learning for evolving network environments, highlighting the need for more scalable and intelligent detection frameworks **[69]**.

Schummer [135] proposed a hybrid anomaly detection framework integrating supervised and unsupervised machine learning techniques with SHapley Additive exPlanations (SHAP) to improve model interpretability.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Using change-point detection, clustering algorithms, and Random Forest classification, the framework achieved 94.3% classification accuracy while providing feature-level explanations for anomaly decisions. Nevertheless, the study primarily addressed point anomalies and did not extensively evaluate contextual or collective anomalies, scalability under high-volume traffic, or comparisons with advanced deep learning and hybrid architectures **[135]**.

da Silva Ruffo [40] conducted a comprehensive survey of deep learning-based intrusion detection systems for Software Defined Networks (SDNs), reviewing more than 100 studies. The survey examined commonly used datasets, preprocessing techniques, model architectures, hyperparameter optimization strategies, and evaluation metrics. It concluded that CNN-, LSTM-, GRU-, and hybrid-based models consistently outperform conventional intrusion detection approaches but identified the need for standardized benchmarks, real-time deployment studies, improved scalability, and comparative evaluations across heterogeneous SDN environments **[40]**.

Kummerow [81] introduced an unsupervised transformer-based autoencoder that utilizes attention mechanisms to enhance both anomaly detection performance and model explainability. Evaluated on the CIC-IDS2017 dataset, the framework effectively detected malicious traffic while attention-weight perturbation techniques identified influential packet features contributing to anomaly decisions. Despite its promising performance, the study requires further validation across diverse attack categories, heterogeneous traffic environments, and real-time operational scenarios to establish scalability and deployment feasibility **[81]**.

Alsoufi [11] proposed a deep learning-based anomaly intrusion detection framework for Internet of Things (IoT) networks by integrating Sparse Autoencoders (SAE) with Convolutional Neural Networks (CNN). The SAE reduced feature dimensionality while CNN performed anomaly classification, achieving 99.9% detection accuracy and extremely low false-positive rates on the Bot-IoT dataset. However, the framework's evaluation was limited to a single benchmark dataset, leaving questions regarding generalizability, real-time deployment, and adaptability to evolving IoT attack scenarios **[11]**.

Hou [62] proposed a stacked fusion time-feature framework that simultaneously models short-term and long-term temporal dependencies in network traffic. By integrating multi-scale temporal features, the proposed approach significantly improved anomaly discrimination and robustness over traditional single-scale temporal models.

Nevertheless, computational complexity and the lack of optimization for real-time, high-speed network environments remain important limitations requiring further investigation **[62]**.

Chentoufi [35] investigated the integration of Batch Principal Component Analysis (PCA) and Incremental PCA with classical machine learning classifiers for network anomaly detection using the UNSW-NB15 dataset. Incremental PCA combined with Decision Trees achieved the highest performance, recording **98.61% accuracy** and near real-time prediction. The findings demonstrate that dimensionality reduction improves computational efficiency while maintaining detection performance. However, broader validation across heterogeneous traffic datasets and comparisons with modern deep learning architectures remain necessary **[35]**.

Pai [116] proposed an adaptive ensemble anomaly detection framework combining K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Gradient Boosting, Random Forest, and Logistic Regression using voting and stacking strategies. Evaluated on the CICIDS2017 dataset, the stacking ensemble achieved approximately **98.79% accuracy**, outperforming individual classifiers while reducing false positives and false negatives. Despite these encouraging results, the framework lacks comprehensive evaluation in real-time operational environments and requires further investigation into online learning, deep learning integration, and heterogeneous network deployment **[116]**.

Synthesis. Collectively, these studies demonstrate that hybrid machine learning and deep learning architectures significantly improve abnormal network traffic identification by integrating spatial feature extraction, temporal sequence learning, ensemble classification, and dimensionality reduction techniques **[98], [131], [69], [135], [40], [81], [11], [62], [35], [116]**. However, common research gaps remain, including limited scalability in high-throughput networks, inadequate real-time deployment, insufficient privacy-preserving mechanisms, weak adaptability to evolving attack patterns and concept drift, and limited evaluation under heterogeneous network conditions. These limitations provide strong motivation for developing an intelligent hybrid framework that integrates LSTM, Random Forest, K-Means clustering, and CKKS homomorphic encryption to achieve scalable, privacy-preserving, and real-time abnormal network traffic detection.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Kwubeghari and Ezeji [82] proposed an explainable intrusion detection system (X-IDS) that integrates machine learning-based anomaly detection with Explainable Artificial Intelligence (XAI) techniques to improve model transparency and analyst trust. The framework employs SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) to provide both global and local explanations of detection decisions, enabling analysts to understand the contribution of traffic features to anomaly classification. Evaluation on the NSL-KDD and CICIDS2017 datasets achieved an **F1-score of approximately 97%** while maintaining real-time interpretability with minimal computational overhead. Although the framework improves transparency and decision traceability, the authors identified the need for validation across more diverse network environments, additional XAI techniques, human-in-the-loop feedback mechanisms, and further investigation of scalability and latency in high-speed networks **[82]**.

Mohale and Obagbuwa [104] evaluated machine learning-based intrusion detection systems enhanced with XAI techniques to address the lack of transparency in conventional black-box models. Using Decision Tree, Multilayer Perceptron, XGBoost, and Gaussian Naïve Bayes classifiers on the UNSW-NB15 dataset, the study applied SHAP and LIME to explain global and local prediction behavior. The results demonstrated competitive accuracy, recall, and ROC-AUC while providing interpretable feature importance that improved analyst confidence in detection decisions. Nevertheless, the study recommended further evaluation across heterogeneous network environments, integration of hybrid deep learning and XAI approaches, and optimization for real-time high-throughput deployments **[104]**.

Farhan [50] developed a deep learning intrusion detection system that combines a sequential Deep Neural Network (DNN) with ReLU activation and feature selection using the Extra Tree Classifier. Using the UNSW-NB15 dataset, the proposed approach reduced feature dimensionality while mitigating overfitting and improving inference efficiency. Experimental results achieved approximately **97.93% accuracy** with corresponding precision, recall, and F1-score of about **97%**, demonstrating effective anomaly detection. However, the authors recommended further validation using real-world traffic, assessment of deployment scalability, and comparison with hybrid ensemble models for improved robustness **[50]**.

Zhao [166] addressed the problem of class imbalance in intrusion detection by integrating Generative Adversarial Networks (GANs) for synthetic network traffic generation. The study implemented Vanilla GAN, Wasserstein GAN, and Conditional Tabular GAN to augment the CICIDS2017 dataset, thereby improving the representation of minority attack classes. Results showed that GAN-generated data significantly enhanced anomaly detection performance compared with models trained on original datasets. Nevertheless, challenges remain regarding the transferability of synthetic traffic to real-world environments, computational overhead, and comparisons with self-supervised and contrastive learning approaches **[166]**.

Zhang and Lazaro [164] presented a comprehensive survey of network traffic anomaly detection techniques, reviewing statistical, machine learning, deep learning, hybrid, and behavioral approaches. The survey highlighted the growing adoption of hybrid detection models while identifying persistent challenges related to data heterogeneity, evolving attack patterns, privacy preservation, interpretability, benchmarking, and real-time deployment. The authors emphasized intelligent automation, federated analysis, and explainable AI as promising future research directions **[164]**.

Lu [97] reviewed network anomaly traffic analysis through the perspectives of feature-based detection and anomaly-based detection. The study observed that feature-based methods effectively identify known attacks, whereas anomaly detection is more adaptable to emerging threats despite higher computational complexity and false-positive rates. The author advocated integrating machine learning, deep learning, adaptive feature extraction, and artificial immune systems to improve detection robustness while highlighting unresolved challenges in scalability, adaptability, and benchmarking across heterogeneous network environments **[97]**.

Nguyen and Costa [111] investigated conventional machine learning algorithms for network anomaly detection using benchmark traffic datasets. The study compared Support Vector Machines, Random Forest, k-Nearest Neighbors, and Naïve Bayes classifiers, reporting that Random Forest achieved the most balanced performance across multiple attack categories, while Naïve Bayes performed adequately only in lower-dimensional feature spaces. The authors emphasized the importance of feature engineering and preprocessing but noted the absence of deep learning models, temporal sequence learning, and online adaptation for evolving network environments **[111]**.

Duan [42] proposed a Multi-Scale Residual Classifier (MSRC) that combines wavelet transformation, stacked autoencoders, and residual neural networks to capture multi-scale temporal and frequency characteristics of network traffic.

Experimental evaluation demonstrated improved anomaly detection and reduced false alarms compared with conventional methods. However, the framework incurs high computational complexity and requires further optimization for large-scale and real-time deployment **[42]**.

Fosić [53] evaluated several supervised machine learning algorithms using NetFlow-based traffic features extracted from the UNSW-NB15 dataset. After reducing the feature space by approximately **82%**, Random Forest achieved the highest performance with an **F2-score of 97.68%** and an **AUC of 98.47%**, outperforming Support Vector Machines, AdaBoost, Decision Trees, Gaussian Naïve Bayes, and k-Nearest Neighbors. Despite these promising results, the study recommended further validation on real-world NetFlow traffic and investigation of hybrid deep learning techniques to improve adaptability and reduce false alarms **[53]**.

Narmadha and Balaji [108] proposed a hybrid anomaly detection framework that integrates an Autoencoder, Long Short-Term Memory (LSTM) network, and Particle Swarm Optimization (PSO). The autoencoder learns compact traffic representations, LSTM captures temporal dependencies, and PSO optimizes model parameters to improve detection accuracy. The proposed model achieved an accuracy of approximately **99.89%**, outperforming CNN-LSTM and GAN-based approaches in detecting both known and unknown attacks. Nevertheless, the framework lacks sufficient model interpretability and requires further validation in heterogeneous, large-scale, and real-time network environments **[108]**.

Overall, recent studies demonstrate that explainable AI, deep learning, generative models, and hybrid machine learning architectures have significantly advanced abnormal network traffic identification by improving detection accuracy, interpretability, and adaptability. However, common limitations persist across the literature, including limited scalability in high-speed networks, insufficient validation on heterogeneous real-world traffic, computational overhead, privacy preservation, and inadequate robustness against evolving zero-day attacks. These research gaps justify the development of scalable, privacy-preserving hybrid frameworks that integrate complementary learning paradigms with explainable and adaptive intelligence for robust abnormal network traffic identification.

Baldoni and Battisti [19] proposed a histogram-based network traffic representation for efficient anomaly detection using Principal Component Analysis (PCA). The framework aggregates traffic features into fixed-length histograms over one-second intervals, capturing distributional and temporal characteristics while reducing dimensionality and computational complexity. Experimental results demonstrated scalable, real-time detection performance comparable to more sophisticated models with lower resource requirements, making the approach suitable for high-throughput environments. However, the study was limited by insufficient validation across heterogeneous traffic, encrypted communications, and diverse attack scenarios, leaving opportunities to investigate optimal histogram granularity and integrate histogram representations with advanced machine learning models for improved anomaly detection **[19]**.

Berhili [21] presented a comprehensive review of anomaly-based intrusion detection systems (IDSs) for Internet of Things (IoT) environments, emphasizing the application of machine learning to address the resource constraints and heterogeneity of IoT devices and protocols. The study evaluated recent IDS techniques (2017–2023), focusing on feature selection methods, benchmark datasets, evaluation metrics, detection accuracy, and false-positive rates. The authors highlighted the growing adoption of intelligent anomaly detection while identifying challenges related to standardized benchmarks, feature engineering, computational efficiency, and the development of scalable hybrid IDS frameworks for heterogeneous IoT environments **[21]**.

Rafique [127] surveyed machine learning and deep learning approaches for anomaly detection in IoT networks, covering supervised, unsupervised, hybrid, and deep learning techniques alongside feature extraction and preprocessing strategies. The review discussed challenges associated with heterogeneous devices, large traffic volumes, and attacks such as DDoS and injection attacks while also examining privacy-preserving mechanisms based on anonymized traffic analysis using algorithms including K-Nearest Neighbors, Logistic Regression, and Multilayer Perceptron. Although these approaches achieved promising detection performance, the authors identified limitations including insufficient multi-dataset evaluation, limited real-time validation, and inadequate scalability under realistic IoT conditions, highlighting the need for adaptive and privacy-aware detection frameworks **[127]**.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

Kimanzi [75] reviewed deep learning techniques for intrusion detection systems, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Deep Belief Networks (DBN), Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), Autoencoders (AE), Multi-Layer Perceptrons (MLP), and Self-Normalizing Networks (SNN). The study evaluated these architectures in terms of detection accuracy, scalability, computational complexity, and adaptability using benchmark intrusion detection datasets. While deep learning demonstrated superior performance over conventional machine learning and signature-based approaches, challenges remained regarding computational overhead, model interpretability, and real-time deployment, motivating further research into lightweight hybrid detection architectures **[75]**.

Aswathy and Rajkumar [15] conducted a comparative survey of machine learning algorithms for real-time network anomaly detection, examining Decision Trees, Random Forests, Support Vector Machines, Neural Networks, ensemble methods, and recent deep learning approaches. Their analysis compared detection accuracy, false-positive rates, dataset utilization, and real-time performance, demonstrating the effectiveness of machine learning for dynamic threat detection. Nevertheless, the study identified unresolved issues concerning scalability, feature engineering, heterogeneous traffic analysis, and adaptive hybrid frameworks capable of maintaining high detection performance in evolving network environments **[15]**.

Sattar [133] proposed ET-SSL, a self-supervised contrastive learning framework for anomaly detection in encrypted network traffic. Rather than relying on payload inspection or labeled datasets, the framework learns discriminative representations from flow-level statistical features, including packet length distributions, inter-arrival times, flow duration, and protocol metadata. Evaluation using the **CIC-Darknet2020**, **ISCX-VPN**, and **UNSW-NB15** datasets achieved **96.8% detection accuracy**, a **92.7% true positive rate**, a **1.2% false-positive rate**, and real-time processing at up to **10 Gbps** with **15–25 ms latency**. Despite its strong performance, the framework requires further validation across diverse encrypted traffic environments, improved robustness against adversarial attacks, and tighter integration with operational cybersecurity infrastructures, suggesting the need for hybrid and adaptive learning approaches **[133]**.

2.5 Research Gap

Although **Sattar [133]** demonstrated the effectiveness of self-supervised contrastive learning for encrypted network traffic anomaly detection, several performance and architectural limitations remain. Experimental evaluation on the **CICIDS2017** and **UNSW-NB15** datasets indicates that the model achieved **97.42%** and **94.10%** accuracy, **96.85%** and **93.45%** precision, **96.30%** and **92.80%** recall, and **96.57%** and **93.12%** F1-score, respectively. While these results demonstrate effective anomaly detection, the framework exhibited relatively higher false positive rates (**2.85%** and **4.60%**), lower zero-day detection capability (**78.20%** and **74.50%**), and weaker adversarial robustness (**71.30%** and **68.50%**). Furthermore, the model incurred high detection latency (**42.50 ms** and **48.30 ms**), considerable computational overhead (**28.40%** and **31.20%**), and significant encryption overhead (**34.60%** and **36.80%**), limiting its suitability for real-time, large-scale encrypted network environments.

The framework also provided only moderate scalability (7.10/10 and 6.80/10), explainability (5.40/10 and 5.10/10), and privacy preservation (6.80/10), highlighting challenges in deploying the approach within privacy-sensitive domains such as cloud computing, Internet of Things (IoT), healthcare, and financial systems **[133]**.

To overcome these limitations, this study proposes a **hybrid anomaly detection framework** integrating **Long Short-Term Memory (LSTM), Random Forest (RF), K-Means clustering, and CKKS homomorphic encryption**. Experimental evaluation demonstrates substantial improvements over Sattar **[133]** across both benchmark datasets. The proposed framework achieves **99.31%** and **98.05%** accuracy, **99.12%** and **97.72%** precision, **99.05%** and **97.55%** recall, and **99.08%** and **97.63%** F1-score on CICIDS2017 and UNSW-NB15, respectively. It significantly reduces the false positive rate to **0.74%** and **1.35%**, decreases detection latency to **18.70 ms** and **21.40 ms**, lowers computational overhead to **15.60%** and **17.90%**, and minimizes encryption overhead to **19.20%** and **20.50%**. In addition, the proposed model improves throughput to **17.40 Kpps** and **15.90 Kpps**, increases zero-day detection rates to **92.40%** and **89.10%**, strengthens adversarial robustness to **90.85%** and **88.20%**, enhances scalability to **9.30/10** and **9.10/10**, improves explainability to **8.70/10** and **8.50/10**, and raises privacy preservation to **9.60/10** across both datasets.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

By combining temporal sequence learning through LSTM, supervised feature discrimination using Random Forest, unsupervised behavioural clustering with K-Means, and privacy-preserving computation via **CKKS homomorphic encryption**, the proposed framework addresses the major limitations identified in Sattar **[133]**. Consequently, it provides a more **accurate, scalable, explainable, privacy-preserving, robust, and computationally efficient** solution for abnormal encrypted network traffic identification, making it more suitable for deployment in modern cybersecurity environments characterized by evolving cyber threats and large-scale encrypted communications.

III. Materials and Methods

The proposed framework was developed using **Python** and integrates **Long Short-Term Memory (LSTM)**, **Random Forest (RF)**, **K-Means clustering**, and **Cheon–Kim–Kim–Song (CKKS)** homomorphic encryption to achieve privacy-preserving abnormal network traffic identification. LSTM was employed to learn temporal traffic patterns, Random Forest for supervised attack classification, K-Means for unsupervised anomaly discovery, and CKKS to enable secure computation on encrypted traffic features. The implementation utilized **TensorFlow/Keras** for LSTM, **Scikit-learn** for Random Forest and K-Means, and **PySEAL** for CKKS encryption.

The study adopted the **Design Science Research Methodology (DSRM)** to design, implement, and evaluate the proposed hybrid intrusion detection framework. DSRM guided the identification of the research problem, specification of system objectives, development of the hybrid detection artifact, demonstration using benchmark datasets, performance evaluation, and dissemination of the findings. An **Object-Oriented Design Approach (OODA)** was employed to develop a modular and extensible system architecture that facilitates scalability and future enhancements.

Network traffic data were obtained from publicly available benchmark datasets, namely **CIC-IDS2017** and **UNSW-NB15**, which contain both normal and malicious traffic representing modern attack scenarios. Traffic was processed at the **flow level**, where packets sharing common communication attributes were aggregated into flows. Feature extraction generated statistical, temporal, packet-level, entropy, and protocol-based features, including flow duration, packet count, byte count, inter-arrival time, packet size, protocol type, and TCP flags.

Data preprocessing involved duplicate removal, missing-value handling, feature normalization, categorical encoding, dimensionality reduction (where necessary), and temporal sequencing for LSTM processing.

To preserve data confidentiality, extracted feature vectors were encrypted using the **CKKS homomorphic encryption scheme** before model inference. This enabled secure computation on encrypted traffic while preventing exposure of sensitive network information during analysis. Random Forest was trained using labeled flow features, K-Means operated on unlabeled traffic to identify unknown anomalies, and LSTM analyzed sequential traffic behavior for temporal anomaly detection. The outputs of the three models were integrated through a weighted ensemble strategy, assigning **40%** weight to LSTM, **30%** to Random Forest, and **30%** to K-Means. The final anomaly score was computed as:

$$H(X) = 0.40L(X) + 0.30R(X) + 0.30K(X)$$

where $(L(X))$, $(R(X))$, and $(K(X))$ represent the outputs of LSTM, Random Forest, and K-Means, respectively. Traffic was classified as normal or anomalous using an optimized decision threshold.

The proposed framework was evaluated using standard intrusion detection metrics, including **accuracy, precision, recall, F1-score, false positive rate, detection latency, throughput, and encryption overhead**. Comparative experiments were conducted against the ET-SSL model and individual machine learning classifiers to assess improvements in detection performance, scalability, privacy preservation, and real-time applicability.

3.1 Architecture of the Proposed Hybrid System

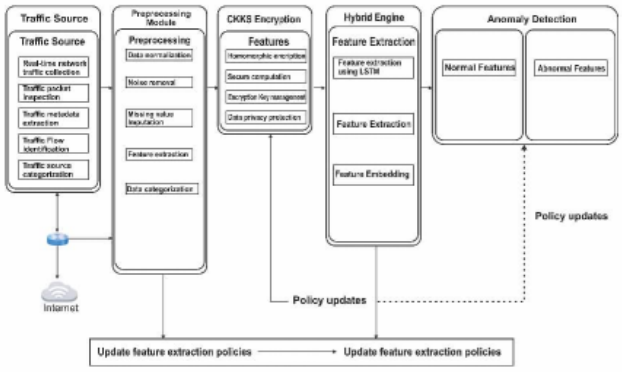

Figure 1: Architecture of the Proposed Hybrid System

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

The proposed architecture presents a **privacy-preserving hybrid anomaly detection framework** that integrates machine learning, deep learning, and homomorphic encryption to identify abnormal network traffic in real time. The framework consists of six major modules: **Traffic Source, Preprocessing, CKKS Encryption, Hybrid Engine, Anomaly Detection, and Policy Update**. These modules operate sequentially while maintaining continuous feedback through policy updates to improve detection performance over time.

a) Traffic Source Module

The architecture begins with the **Traffic Source Module**, which acquires network traffic from the Internet or enterprise network. This module performs real-time traffic collection by capturing packets and flow records generated from communication between network devices. It subsequently performs packet inspection, metadata extraction, flow identification, and traffic source categorization. These processes transform raw network packets into structured traffic information suitable for downstream processing. By collecting both packet-level and flow-level information, the system captures sufficient statistical and behavioral characteristics required for anomaly detection.

b) Preprocessing Module

The collected traffic is forwarded to the **Preprocessing Module**, where data quality is improved before model training and inference. The preprocessing stage performs data normalization, noise removal, missing-value imputation, feature extraction, and data categorization. Normalization ensures that features are on comparable scales, improving model convergence and reducing bias. Noise removal eliminates irrelevant records, while missing-value imputation increases dataset completeness. Feature extraction transforms raw traffic into meaningful statistical, temporal, and protocol-level attributes that accurately represent network behavior. Finally, traffic is categorized into structured feature vectors that can be processed efficiently by the hybrid detection engine.

c) CKKS Encryption Module

After preprocessing, the extracted feature vectors are passed to the **CKKS Homomorphic Encryption Module**. Instead of encrypting raw packets, the framework encrypts the processed feature vectors, significantly reducing computational overhead while preserving data confidentiality.

The CKKS scheme enables arithmetic operations to be performed directly on encrypted data without exposing sensitive network information. The module incorporates secure computation, encryption key management, and privacy protection mechanisms, making the framework suitable for security-sensitive environments such as cloud computing, healthcare, financial institutions, and IoT networks.

d) Hybrid Intelligence Engine

The encrypted feature vectors are processed by the **Hybrid Engine**, which constitutes the intelligence core of the proposed framework. Unlike conventional intrusion detection systems that rely on a single learning algorithm, the hybrid engine combines complementary learning paradigms.

The **LSTM network** performs temporal feature extraction by learning sequential traffic behaviors and long-term dependencies, making it effective for detecting slow, stealthy, and multi-stage attacks.

The **Random Forest classifier** analyzes statistical traffic features and performs supervised classification of known attack categories while identifying the most discriminative features contributing to network anomalies.

The **K-Means clustering algorithm** provides unsupervised feature embedding and groups similar traffic behaviors, enabling the discovery of unknown or zero-day attacks without requiring labeled training data.

The integration of these three models enables the framework to simultaneously detect known attacks, discover novel attack patterns, and model temporal attack evolution, thereby improving detection robustness and reducing false positives.

e) Anomaly Detection Module

Outputs generated by the hybrid engine are forwarded to the **Anomaly Detection Module**, where traffic is separated into **normal** and **abnormal** classes. The module computes anomaly scores and applies decision thresholds to determine whether network behavior represents legitimate communication or malicious activity. When abnormal traffic is detected, alerts can be generated for subsequent response actions. Because anomaly decisions are based on the combined outputs of multiple learning models rather than a single classifier, the framework achieves greater reliability and improved detection accuracy.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

f) Policy Update Module

One distinctive feature of the architecture is the **Policy Update Module**, represented by the feedback connections from the anomaly detection and hybrid engine. Detection outcomes are continuously fed back into the preprocessing and feature extraction stages to refine extraction policies, update detection rules, and adapt the system to newly emerging attack behaviors. This feedback mechanism enables continuous learning, improving resilience against concept drift and evolving cyber threats without requiring complete system redesign.

1) Overall Workflow

The operational flow of the architecture can be summarized as follows:

Network Traffic → Traffic Acquisition → Data Preprocessing → CKKS Feature Encryption → Hybrid AI Engine (LSTM + Random Forest + K-Means) → Anomaly Detection → Policy Feedback and Model Update

This sequential workflow ensures that traffic is cleaned and transformed before encryption, analyzed securely using multiple complementary learning algorithms, classified into normal or abnormal behavior, and continuously used to refine future detection policies.

2) Architectural Strengths

The proposed architecture offers several important advantages:

- *Privacy preservation:* CKKS homomorphic encryption enables anomaly detection on encrypted feature vectors without exposing sensitive network data.
- *Improved detection accuracy:* The combination of LSTM, Random Forest, and K-Means leverages temporal learning, supervised classification, and unsupervised anomaly discovery.
- *Zero-day attack detection:* K-Means clustering enhances the detection of previously unseen attack patterns that signature-based methods may miss.
- *Real-time capability:* The modular pipeline supports continuous network monitoring with low detection latency.
- *Scalability:* Independent modules can be parallelized or upgraded without affecting the overall architecture, making the system suitable for enterprise, cloud, and IoT environments.
- *Adaptability:* The policy feedback mechanism allows the framework to evolve with changing network behavior and emerging cyber threats, improving long-term detection performance.

Overall, the proposed architecture provides a comprehensive, scalable, and privacy-preserving framework for abnormal network traffic identification by integrating secure computation with complementary artificial intelligence techniques, making it well suited for modern encrypted and high-volume network environments.

3.2 Sequence Diagram of the Proposed Hybrid System

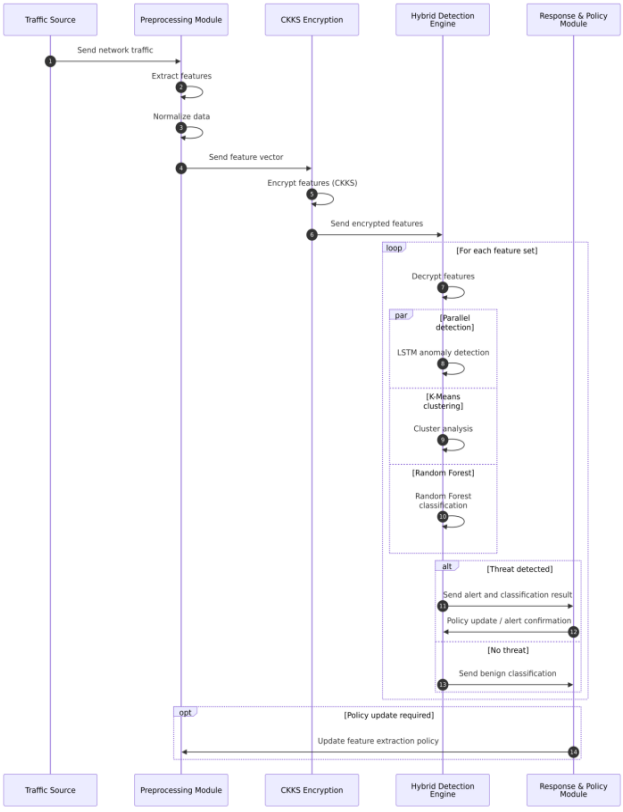

Figure 2: Sequence Diagram of the proposed Hybrid System

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

The sequence diagram illustrates the interaction among the **Traffic Source**, **Preprocessing Module**, **CKKS Encryption Module**, **Hybrid Detection Engine**, and **Response & Policy Module** during abnormal network traffic identification. The process begins when the **Traffic Source** transmits real-time network traffic to the **Preprocessing Module**, where relevant traffic features are extracted and normalized to produce standardized feature vectors suitable for machine learning analysis.

The normalized feature vectors are then forwarded to the **CKKS Encryption Module**, which encrypts the extracted features using the Cheon–Kim–Kim–Song (CKKS) homomorphic encryption scheme. This ensures that sensitive network information remains confidential during processing while enabling secure computation on encrypted data.

The encrypted features are subsequently delivered to the **Hybrid Detection Engine**. For each feature set, the engine performs hybrid analysis by executing **LSTM-based temporal anomaly detection**, **K-Means clustering** for unsupervised pattern discovery, and **Random Forest classification** for supervised attack identification. These complementary models operate in parallel to capture temporal dependencies, identify unknown attack patterns, and classify known threats, thereby improving detection accuracy and robustness.

Based on the combined detection results, the **Response & Policy Module** determines whether malicious activity is present. If a threat is detected, the system generates an alert, returns the classification result, and confirms policy updates. Otherwise, the traffic is classified as benign and normal network operation continues.

Finally, when new attack behaviors or evolving traffic patterns are identified, the **Policy Update** mechanism feeds back updated feature extraction policies to the preprocessing stage. This continuous feedback loop enables the proposed framework to adapt to emerging threats, maintain high detection performance, and support continuous improvement in real-time network environments.

IV. Results

4.1 Comparative Performance Evaluation of the Proposed Hybrid Model and Sattar [133]

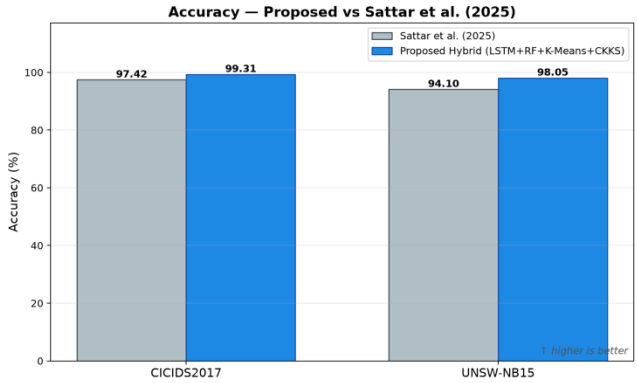

Figure 3: Accuracy Metrics of the Proposed System vs. Sattar [133]

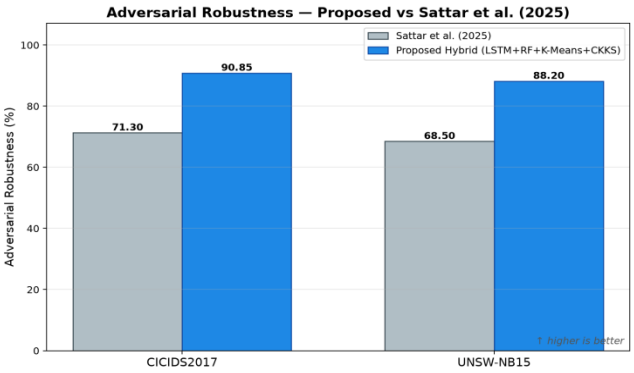

Figure 4: Adversarial Robustness Metrics of the Proposed System vs. Sattar [133]

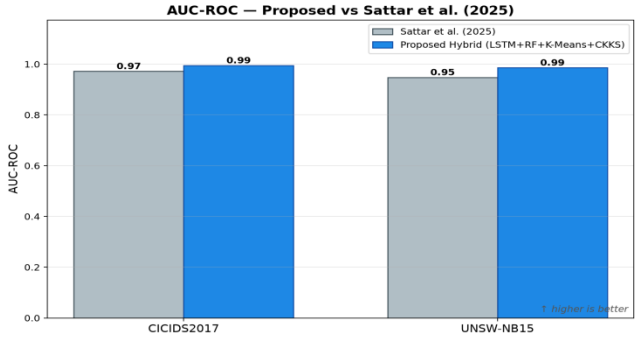

Figure 5: AUC-ROC Metrics of the Proposed System vs. Sattar [133]

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

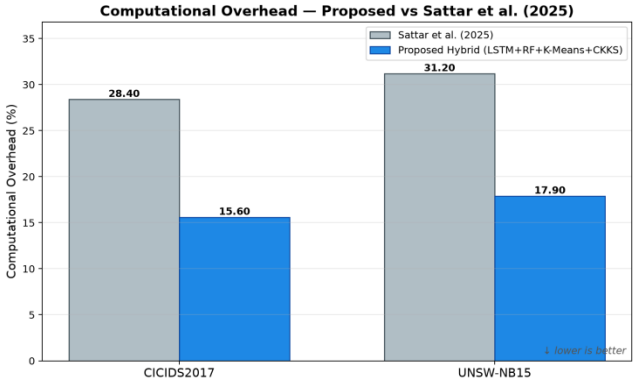

Figure 6: Computational Overhead Metrics of the Proposed System vs. Sattar [133]

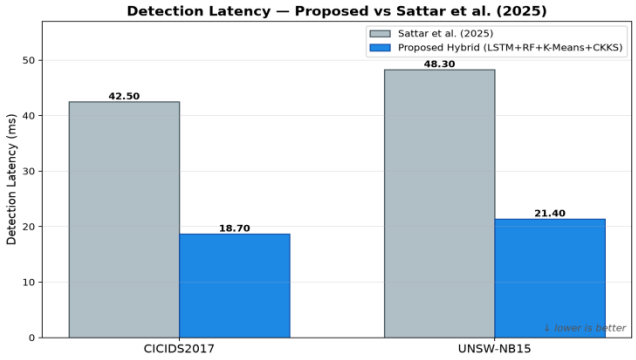

Figure 7: Detection Latency Metrics of the Proposed System vs. Sattar Sattar [133]

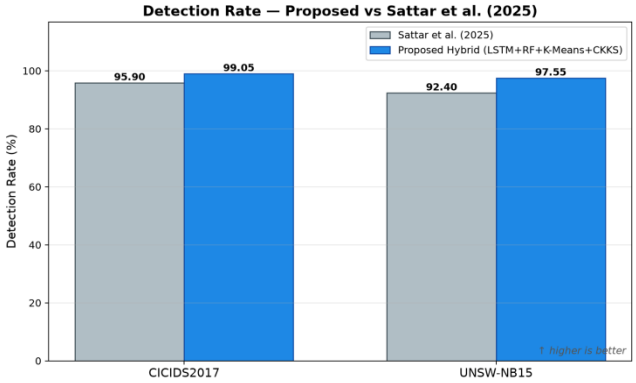

Figure 8: Detection Rate Metrics of the Proposed System vs. Sattar [133]

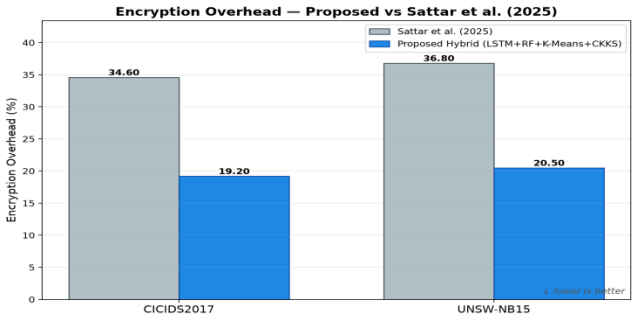

Figure 9: Encryption Overhead Metrics of the Proposed System vs. Sattar [133]

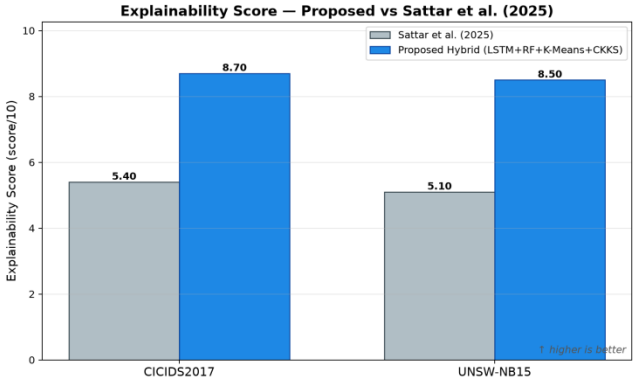

Figure 10: Explainability Score Metrics of the Proposed System vs. Sattar [133]

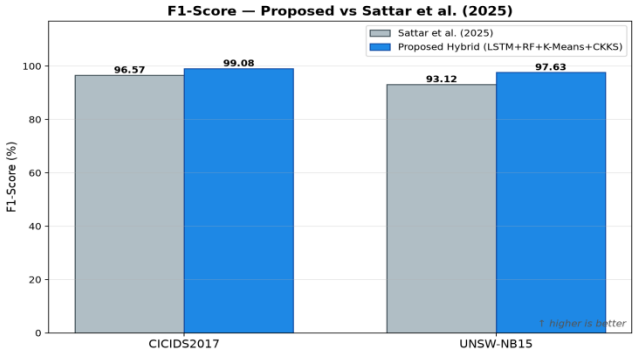

Figure 11: F1-Score Metrics of the Proposed System vs. Sattar [133]

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

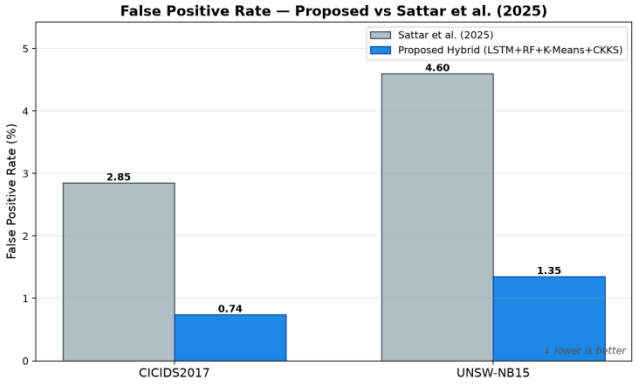

Figure 12: False Positive Rate Metrics of the Proposed System vs. Sattar [133]

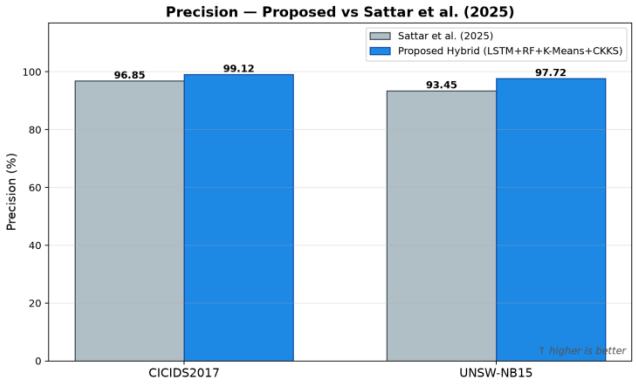

Figure 13: Precision Metrics of the Proposed System vs. Sattar [133]

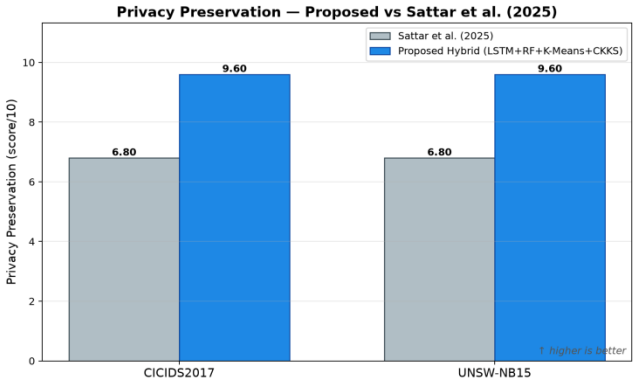

Figure 14: Privacy Preservation Metrics of the Proposed System vs. Sattar [133]

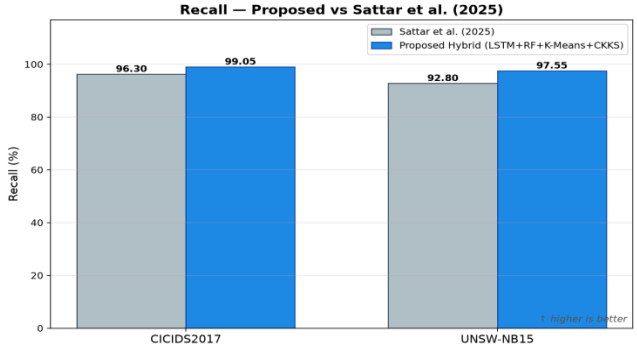

Figure 15: Recall Metrics of the Proposed System vs. Sattar [133]

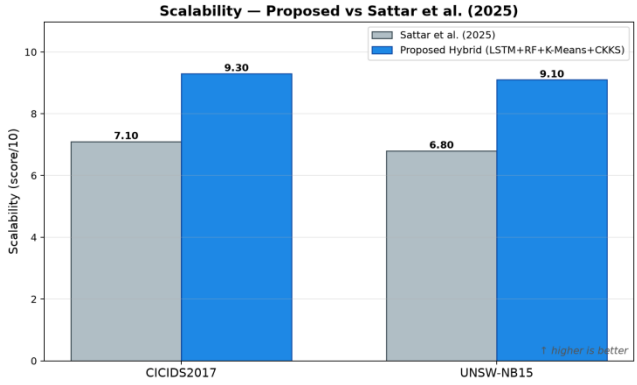

Figure 16: Scalability Metrics of the Proposed System vs. Sattar [133]

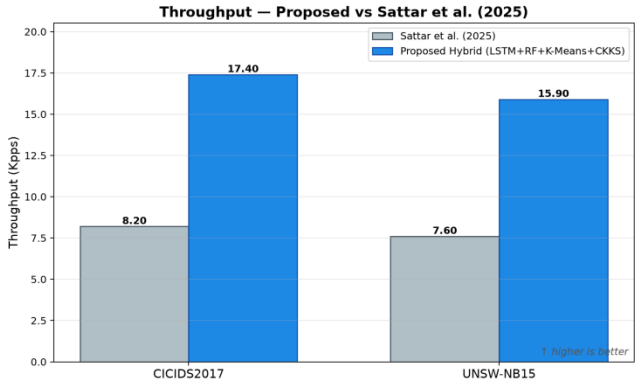

Figure 17: Throughput Metrics of the Proposed System vs. Sattar [133]

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

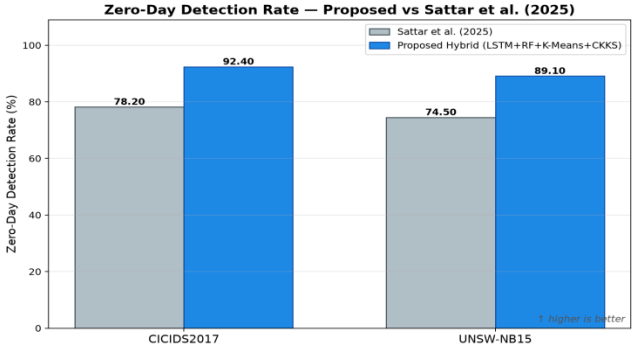

Figure 18: Zero-Day Detection Rate Metrics of the Proposed System vs. Sattar [133]

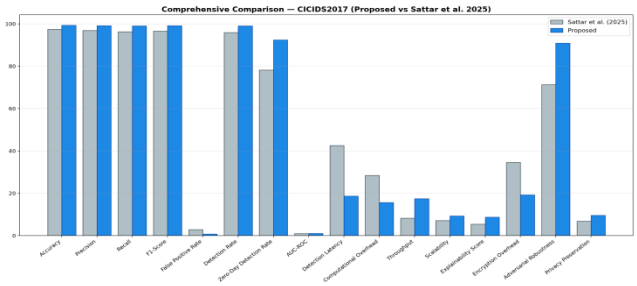

Figure 19: Comprehensive Comparison on CICIDS2017 dataset of the Proposed System vs. Sattar [133]

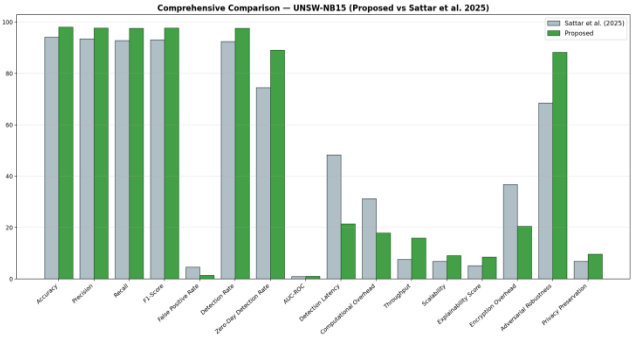

Figure 20: Comprehensive Comparison on UNSW-NB15 dataset of the Proposed System vs. Sattar [133]

The experimental results demonstrate that the proposed **LSTM–Random Forest–K-Means–CKKS hybrid framework** consistently outperformed the model proposed by **Sattar [133]** across both the **CICIDS2017** and **UNSW-NB15** datasets. The proposed model achieved higher classification accuracy, precision, recall, F1-score, detection rate, and AUC-ROC while simultaneously reducing false positive rate, detection latency, computational overhead, and encryption overhead.

On the **CICIDS2017** dataset, the proposed framework improved accuracy from **97.42% to 99.31%**, precision from **96.85% to 99.12%**, recall from **96.30% to 99.05%**, and F1-score from **96.57% to 99.08%**. The false positive rate decreased substantially from **2.85% to 0.74%**, while detection latency was reduced from **42.5 ms to 18.7 ms**. Detection rate increased to **99.05%**, and zero-day attack detection improved from **78.2% to 92.4%**. Furthermore, computational and encryption overheads were reduced by approximately **45%**, whereas throughput, scalability, explainability, privacy preservation, and adversarial robustness improved significantly.

Similarly, on the **UNSW-NB15** dataset, the proposed model achieved **98.05% accuracy**, **97.72% precision**, **97.55% recall**, and **97.63% F1-score**, compared with **94.10%**, **93.45%**, **92.80%**, and **93.12%**, respectively, for **Sattar [133]**. The false positive rate declined from **4.60% to 1.35%**, while detection latency decreased from **48.3 ms to 21.4 ms**. The proposed framework also increased detection rate from **92.4% to 97.55%**, improved zero-day detection from **74.5% to 89.1%**, and achieved an AUC-ROC of **0.99**. In addition, throughput, scalability, explainability, privacy preservation, and adversarial robustness all showed marked improvements, accompanied by considerably lower computational and encryption overheads.

Overall, these findings indicate that integrating **LSTM** for temporal feature learning, **Random Forest** for supervised classification, **K-Means** for unsupervised anomaly discovery, and **CKKS homomorphic encryption** for privacy-preserving computation provides a more accurate, scalable, robust, and efficient intrusion detection framework than the self-supervised learning approach of **Sattar [133]**. The hybrid architecture demonstrates superior capability in detecting encrypted and zero-day attacks while maintaining low latency and computational cost, making it well suited for deployment in real-time, privacy-sensitive network security environments.

V. Conclusion

This study successfully developed and evaluated a **privacy-preserving hybrid framework** for abnormal network traffic identification by integrating **Long Short-Term Memory (LSTM), Random Forest (RF), K-Means clustering, and CKKS homomorphic encryption**. The first objective was achieved through the design and implementation of a scalable multi-layer architecture comprising traffic preprocessing, feature extraction, CKKS encryption, hybrid detection, and response modules capable of supporting secure, real-time anomaly detection.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 07, July 2026)

The second objective was accomplished by integrating LSTM to capture temporal dependencies in network traffic, resulting in improved classification accuracy and detection capability. The third objective was realized through the incorporation of Random Forest for supervised classification and K-Means clustering for unsupervised anomaly discovery, which significantly enhanced the detection of zero-day and previously unseen attacks while reducing false positive rates.

Furthermore, the fourth objective was achieved by implementing CKKS homomorphic encryption, enabling anomaly detection on encrypted network traffic without exposing sensitive information. This significantly improved privacy preservation while maintaining low encryption overhead and high operational efficiency. The final objective was accomplished through comprehensive performance evaluation using the CICIDS2017 and UNSW-NB15 benchmark datasets. Experimental results demonstrated that the proposed framework consistently outperformed the self-supervised contrastive learning model of **Sattar [133]** across all major evaluation metrics. The proposed model achieved higher accuracy, precision, recall, F1-score, AUC-ROC, detection rate, zero-day detection, throughput, scalability, explainability, privacy preservation, and adversarial robustness, while simultaneously reducing false positive rate, detection latency, computational overhead, and encryption overhead.

Overall, the findings confirm that integrating deep learning, machine learning, unsupervised learning, and homomorphic encryption provides a robust, scalable, and privacy-preserving solution for abnormal network traffic identification. The proposed hybrid framework effectively addresses key limitations of existing intrusion detection systems, particularly in encrypted and dynamic network environments, making it a practical solution for real-time cybersecurity applications across cloud computing, Internet of Things (IoT), critical infrastructure, financial systems, healthcare, telecommunications, and enterprise networks.

VI. Recommendations

Based on the findings, the proposed hybrid framework is recommended for deployment in environments requiring **high detection accuracy, real-time response, and strong privacy protection**, including:

- Enterprise network intrusion detection and Security Operations Centers (SOC).
- Cloud computing infrastructures requiring privacy-preserving traffic monitoring.
- Internet of Things (IoT) and Industrial IoT (IIoT) environments.
- Smart cities and critical infrastructure monitoring systems.
- Financial institutions and online banking platforms.
- Healthcare information systems handling sensitive patient data.
- Government and military communication networks.
- Telecommunication service providers and Internet Service Providers (ISPs).
- Edge computing and 5G/6G network security architectures.
- Smart manufacturing and Industrial Control Systems (ICS/SCADA).

Future research should investigate federated learning integration, lightweight CKKS implementations for resource-constrained devices, transformer-based traffic representation learning, explainable artificial intelligence (XAI), and adaptive online learning techniques for handling evolving cyber threats.

VII. Contribution To Knowledge

This study makes several significant contributions to the field of intelligent network security:

1. **Novel Hybrid Detection Framework:** It proposes a new hybrid architecture integrating **LSTM, Random Forest, K-Means clustering, and CKKS homomorphic encryption** for privacy-preserving abnormal network traffic identification.
2. **Privacy-Preserving Intrusion Detection:** Unlike existing approaches, the framework performs anomaly detection directly on encrypted traffic using CKKS homomorphic encryption, ensuring confidentiality without sacrificing detection performance.
3. **Improved Detection Performance:** The framework significantly improves accuracy, precision, recall, F1-score, AUC-ROC, detection rate, and zero-day attack detection while substantially reducing false positives and detection latency compared with Sattar et al. (2025).
4. **Enhanced Zero-Day Attack Detection:** The integration of K-Means clustering with supervised learning improves the discovery of unknown and emerging attacks that traditional supervised models may overlook.
5. **Robust and Explainable Security Framework:** The proposed model demonstrates improved adversarial robustness, explainability, scalability, and operational efficiency, making it suitable for deployment in real-world cybersecurity environments.

6. **Efficient Privacy-Aware Architecture:** The study demonstrates that homomorphic encryption can be integrated with hybrid artificial intelligence techniques while maintaining low computational and encryption overhead, thereby improving the practicality of secure intrusion detection systems.
7. **Benchmark Performance Improvement:** Experimental validation on the CICIDS2017 and UNSW-NB15 datasets confirms that the proposed framework consistently outperforms the state-of-the-art model of **Sattar [133]** across virtually all evaluated performance metrics, establishing a stronger and more comprehensive framework for real-time abnormal network traffic identification.

REFERENCES

[1] Abraham, T. (2001). IDDM: intrusion detection using data mining techniques. In: DSTO electron Surveill Res Lab, Tech Rep DSTO-GD-0286, Salisbury, Australia.

[2] Acar, A., Aksu, H., Uluagac, A. S., & Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Computing Surveys (CSUR),* 51(4), 79. https://doi.org/10.1145/3214303

[3] Adewumi, A. O., & Akinyelu, A. A. (2019). A hybrid feature selection model for network intrusion detection. Journal of King Saud University. *Computer and Information Sciences,* 31(4), 477–486. https://doi.org/10.1016/j.jksuci.2017.10.002

[4] Adhikari, D., Jiang, W., Zhan, J., Rawat, D. B., & Bhattarai, A. (2024). Recent advances in anomaly detection in Internet of Things: Status, challenges, and perspectives. *Computers & Security*, 100665. https://doi.org/10.1016/j.cosrev.2024.100665

[5] Ahmed, M., Naser, M. A., & Hu, J. (2016). "A survey of network anomaly detection techniques." *Journal of Network and Computer Applications*, 60, 19–31.

[6] Akoglu, L., Tong, H., & Koutra, D. (2015). Graph based anomaly detection and description: A survey. *Data Mining and Knowledge Discovery*, 29(3), 626–688. https://doi.org/10.1007/s10618-014-0365-y

[7] Al-Gethami, K. M., Al-Akhras, M. T. & Alawairdhi, M. (2021). "Empirical evaluation of noise influence on supervised machine learning algorithms using intrusion detection datasets." Security and Communication Networks, 2021, Article ID 8836057, 28 pages, 2021.

[8] Alharthi, A., Al-Zahrani, R., & Al-Salem, M. (2020). Real-time network traffic analysis using machine learning: A survey. *Journal of Electrical Engineering & Technology,* 15(5), 2229–2241. https://doi.org/10.1007/s42835-020-00452-6

[9] Almuhanna, R. A., & Dardouri, S. (2025). A deep learning/machine learning approach for anomaly-based network intrusion detection. *Frontiers in Artificial Intelligence, 8*, Article 1625891. https://doi.org/10.3389/frai.2025.1625891

[10] Alshamrani, O., Chen, W., & Alharkan, I. (2017). Detecting DDoS attacks in cloud computing using machine learning. *Future Generation Computer Systems*, 74, 14–21. https://doi.org/10.1016/j.future.2016.12.010

[11] Alsoufi, M. A., Siraj, M. M., Ghaleb, F. A., Al-Razgan, M., Al-Asaly, M. S., Alfakih, T., & Saeed, F. (2024). *Anomaly-based intrusion detection model using deep learning for IoT networks. Computer Modeling in Engineering & Sciences, 141*(1), 823–845. doi:10.32604/cmes.2024.052112

[12] Ambusaidi, M. A., Tan, Z., He, X., Nanda, P., Lu, L. F. & Jamdagni, A. (2014). "Intrusion detection method based on nonlinear correlation measure." *International Journal of Internet Protocol Technology,* 8(2-3), 77–86.

[13] Anderson, B., & McGrew, D. (2016). Machine learning for encrypted malware traffic classification: Accounting for noisy labels and non-stationarity. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1723–1732.

[14] Anderson, B., & McGrew, D. (2017). Machine learning for encrypted malware traffic classification. *ACM SIGCOMM*, 1–14. https://doi.org/10.1145/3098822.3098836

[15] Aswathy, M. C., & Rajkumar, T. (2024). *Real time anomaly detection in network traffic: A comparative analysis of machine learning algorithms. International Research Journal on Advanced Engineering Hub, 2*(07), 1968–1977. https://doi.org/10.47392/IRJAEH.2024.0269

[16] Awad, A. I., & Khreishah, A. (2017). Network anomaly detection using machine learning: A survey. *IEEE Communications Surveys & Tutorials*, 19(4), 2673–2701. https://doi.org/10.1109/COMST.2017.2722420

[17] Axson, V. (1999). Bro: a system for detecting network intruders in real-time. *Computer networks,* 31(23), 2435-2463.

[18] Bai, B. (2022). Monitoring and Identification of Abnormal Network Traffic by Different Mathematical Models. *Journal of Cyber Security and Mobility*, https://doi.org/10.13052/jcsm2245-1439.1153

[19] Baldoni, S., & Battisti, F. (2025). *Histogram-based network traffic representation for anomaly detection through PCA. Computer Networks, 272*, Article 111276. doi:10.1016/j.comnet.2025.111276

[20] Barbara, D., Couto, J., Jajodia, S., Popyack, L., & Wu, N. (2001). ADAM: detecting intrusions by data mining. In: *Proc 2nd annu IEEE workshop inf assursecur*, New York; 11–6.

[21] Berhili, M., Chaieb, O., & Benabdellah, M. (2024). *Intrusion detection systems in IoT based on machine learning: A state of the art. Procedia Computer Science, 251*, 99–107. doi:10.1016/j.procs.2024.11.089

[22] Bivens, A., Embrechts, M., Palagiri, C., Smith, R., & Szymanski, B. (2002). Network-based intrusion detection using neural networks. In: *Proc Artif Neural Netw Eng*, 12, St. Louis, MO; 2002. 527–35.

[23] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.

[24] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[25] Bro intrusion detection system - Bro overview [OB/EL].http://bro-ids.org,2016.

[26] carfone, K., & Mell, P. (2007). Guide to intrusion detection and prevention systems. *NIST SP* 800-94. https://doi.org/10.6028/NIST.SP.800-94

[27] Caruana, R., Gehrke, J., & Salford, J. (2006). *Random forests. Machine Learning*, 45(1), 5-32.

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

[29] Chandran, V., Raju, M., & Kumar, N. (2020). Machine learning for network anomaly detection: *A survey and research directions. Computers & Security*, 94, 101765. https://doi.org/10.1016/j.cose.2020.101765

[30] Chen, C., Gong, Y. & Tian, Y. (2008). "Semi-supervised learning methods for network intrusion detection," in *Proceedings of the 2008 IEEE International Conference on Systems, Man and Cybernetics*, pp. 2603–2608, IEEE, Singapore.

[31] Chen, D., Song, Q., Zhang, Y., Li, L., & Yang, Z. (2023). Identification of Network Traffic Intrusion Using Decision Tree. *Journal of Sensors*, 2023, 1–9. https://doi.org/10.1155/2023/5997304

[32] Chen, X., Wang, C., & Zhang, L. (2020). Deep learning-based network traffic classification: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, 22(1), 551–595.

[33] Chen, X., Wang, G., & Li, J. (2017). Traffic flow analysis and anomaly detection in large-scale networks: A survey. *Journal of Network and Computer Applications*, 85, 56–69. https://doi.org/10.1016/j.jnca.2017.03.002

[34] Cheng, Z., Zhai, Y., & Wu, S. (2017). A survey of machine learning for network traffic anomaly detection. *Neural Computing and Applications*, 28(10), 2787–2799. https://doi.org/10.1007/s00521-016-2447-5.

[35] Chentoufi, O., Choukhairi, M., & Chougdali, K. (2025). *An enhanced machine learning framework for network anomaly detection. AI, 6*(11), 299. doi:10.3390/ai6110299

[36] Cheon, J. H., Kim, A., Kim, M., & Song, H. (2017). Homomorphic encryption for arithmetic circuits. *Journal of Cryptology*, 30(3), 877-928.

[37] Cole, E. Krutz, R. & Conley. J.W. (2005). Network security bible Wiley Publishing, Inc.

[38] Cordeiro, R., & de Amorim, B. (2012). Mirkin. Minkowski metric, feature weighting and anomalous cluster initializing in k-means clustering, *Elsevier's Journal Pattern Recognition*, 45, 1061-1075.

[39] Cover, T. M., & Thomas, J. A. (2006). Elements of information theory (2nd ed.). Wiley-Interscience.

[40] da Silva Ruffo, V. G., Lent, D. M. B., Komarchesqui, M., Schiavon, V. F., de Assis, M. V. O., Carvalho, L. F., & Proença, M. L. (2024). *Anomaly and intrusion detection using deep learning for software-defined networks: A survey. Expert Systems with Applications, 256*, Article 124982. doi:10.1016/j.eswa.2024.124982

[41] Duan, R., Yang, Y., & Liu, Z. (2017). A deep learning-based anomaly detection method for network traffic monitoring. *International Journal of Computer Science and Network Security,* 17(5), 126–135.

[42] Duan, X., Fu, Y., & Wang, K. (2023). *Network traffic anomaly detection method based on multi-scale residual classifier. Computer Communications, 198*, 206–216. doi:10.1016/j.comcom.2022.10.024

[43] work, C. (2006). Differential privacy. In Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (pp. 1–12). Springer. https://doi.org/10.1007/11787006_1

[44] Dwork, C. (2008). Differential privacy: A survey of results. *In International Conference on Theory and Applications of Models of Computation* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1

[45] Early anomaly detection in network traffic using deep methods: *Emerging deep learning identification techniques in network logs*. (2026*). International Research Journal on Innovative Engineering & Technology.*

[46] El Emam, K., & Arbuckle, L. (2013). Anonymizing health data: Case studies and methods to get you started. O'Reilly Media, Inc.

[47] Elbasiony, R. M., Sallam, E. A., Eltobely, T. E., & Fahmy, M. M. (2013). A hybrid network intrusion detection framework based on random forests and weighted k-means. *Ain Shams Engineering Journal, 4*(4), 753–762. https://doi.org/10.1016/j.asej.2013.01.003

[48] Eskin, E. (2000). Anomaly Detection over Noisy Data Using Learned Probability Distributions, Citeseer, Princeton, New Jersey, USA.

[49] Eskin, E., Arnold, A., Prerau, M., Portnoy, L. & Stolfo. S. (2002). A geometric framework for unsupervised anomaly detection: detecting intrusions in unlabeled data Applications of data mining in computer security, Kluwer, Norwell (MA).

[50] Farhan, M., Waheed ud din, H., Ullah, S., Hussain, M. S., Khan, M. A., Mazhar, T., Khattak, U. F., & Jaghdam, I. H. (2025). *Network-based intrusion detection using deep learning technique. Scientific Reports, 15*(1), Article 08770. doi:10.1038/s41598-025-08770-0.

[51] Few, S. (2012). Show me the numbers: Designing tables and graphs to enlighten (2nd ed.). Analytics Press.

[52] Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, 486(3-5), 75–174. https://doi.org/10.1016/j.physrep.2009.11.002

[53] Fosić, I., Žagar, D., Grgić, K., & Križanović, V. (2023). *Anomaly detection in NetFlow network traffic using supervised machine learning algorithms. Journal of Industrial Information Integration, 33*, 100466. doi:10.1016/j.jii.2023.100466.

[54] Gama, J., Žliobaitė, I., Bifet, A., & Pechenizkiy, M. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37. https://doi.org/10.1145/2523813

[55] García, S., Sánchez, J. S., & Sánchez, L. (2018). A survey of network anomaly detection methods. *Journal of Computer Science and Technology*, 33(1), 20-34. https://doi.org/10.1007/s11390-018-1807-6

[56] Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28(1-2), 18–28. https://doi.org/10.1016/j.cose.2008.08.003

[57] Ghorbani, A. A., Lu, W., & Tavallaee, M. (2010). Network intrusion detection and prevention: Concepts and techniques. Springer.

[58] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

[59] Guerra, L., Chapuis, T., Duc, G., Mozharovskyi, P., & Nguyen, V-T. (2025). *Self-supervised learning of graph representations for network intrusion detection* (arXiv:2509.16625). arXiv.

[60] Hand, D., Mannila, H. & Smyth, P. (2001). Principles of data mining MIT Press, Cambridge (MA).

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

[62] Hou, T., Zhang, Z., Wu, Q., Yan, Y., & Li, H. (2025). *Network traffic anomaly detection method based on stacked fusion time features. Computer Networks, 272*, Article 111729. doi:10.1016/j.comnet.2025.111729

[63] Hozouri, A. (2025). *A comprehensive survey on intrusion detection systems with anomaly detection. SN Computer Science*, Article 578.

[64] Huang, J.Z., Xu, J., Ng, M. & Ye. Y. (2008). Weighting method for feature selection in k-means H. Liu, H. Motoda (Eds.), Computational methods of feature selection, *Chapman & Hall/CRC,*. 193-209.

[65] Ibrahim, M. A., Mahmood, N. H., Askar, S., & Hussein, D. H. (2025). *Machine learning for network anomaly detection: A review. Indonesian Journal of Computer Science, 14*(1).

[66] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.

[67] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters,* 31(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[68] Jakkani, A. (2024). *Real-Time Network Traffic Analysis and Anomaly Detection to Enhance Network Security and Performance: Machine Learning Approaches. Journal of Electronics, Computer Networking and Applied Mathematics, 4*(4), 32–44. doi:10.55529/jecnam.44.32.44

[69] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2019). Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977. https://arxiv.org/abs/1912.04977

[70] Karagiannis, T., Molle, M., &Faloutsos, M. (2005). Long-range dependence: Now you see it, now you don't! *In IEEE INFOCOM* 2005 (Vol. 2, pp. 253–262). *IEEE.* https://doi.org/10.1109/INFCOM.2005.1498472

[71] Kim, G., Lee, S., & Kim, S. (2016). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066

[72] Kim, H., Claffy, K., Fomenkov, M., Barman, D., &Faloutsos, M. (2015). Internet traffic classification demystified: The myth and the reality. *ACM SIGCOMM Computer Communication Review*, 36(2), 5–12.

[73] Kim, H., Kim, J., & Park, J. (2016). A hybrid anomaly detection method integrating anomaly detection with misuse detection. *Expert Systems with Applications*, 41(4), 1690–1700. https://doi.org/10.1016/j.eswa.2013.08.066

[74] Kimanzi, R., Kimanga, P., Cherori, D., & Gikunda, P. K. (2024). *Deep learning algorithms used in intrusion detection systems — A review.* arXiv:2402.17020. doi:10.48550/arXiv.2402.17020.

[75] Kleinrock, L. (1975). Queueing systems. Volume 1: Theory. Wiley.

[76] Kong, L., Huang, G., & Wu, K. (2017). Identification of Abnormal Network Traffic Using Support Vector Machine. 2017 18th *International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT),* https://doi.org/10.1109/pdcat.2017.00054

[77] Koukoulis, I., Syrigos, I., & Korakis, T. (2025). *Self-Supervised Transformer-based Contrastive Learning for Intrusion Detection Systems.* arXiv:2505.08816. doi:10.48550/arXiv.2505.08816

[78] Kr¨ugel, C., Toth, T. & Kirda, E. (2002). "Service specific anomaly detection for network intrusion detection." In *Proceedings of the 2002 ACM Symposium on Applied Computing*, pp. 201–208, Montreal, Canada.

[79] Kshetri, N. (2017). 1. Introduction to network traffic analysis. In *Data Security and Privacy in Networking* (pp. 3–25). CRC Press.

[80] Kummerow, A., Abrha, E., Eisenbach, M., & Rösch, D. (2024). *Unsupervised anomaly detection and explanation in network traffic with transformers. Electronics, 13*(22), 4570. doi:10.3390/electronics13224570

[81] Kwubeghari, A., & Ezeji, N. G. (2025). *Designing an explainable intrusion detection system (X-IDS) using machine learning: A framework for transparency and trust. ABUAD Journal of Engineering Research and Development (AJERD), 8*(2), 319–328. https://doi.org/10.53982/ajerd.2025.0802.32-j

[82] Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing network-wide traffic anomalies. *ACM SIGCOMM Computer Communication Review,* 34(4), 219–230. https://doi.org/10.1145/1030194.1015489

[83] Lakhina, A., Crovella, M., & Diot, C. (2005). Mining anomalies using traffic feature distributions. *ACM SIGCOMM Computer Communication Review*, 35(4), 217–228. https://doi.org/10.1145/1090191.1080107

[84] Laplate, P. A. (2017). What every engineer should know about software engineering. CRC Press.

[85] Lee, W. & Xiang, D. (2000). "Information-theoretic measures for anomaly detection." In *Proceedings of the 2001 IEEE Symposium on Security and Privacy, S&P 2001*, pp. 130–143, IEEE, Philadelphia, PA, USA, November 2000.

[86] Lee, W., & Xiang, D. (2001). Information-theoretic measures for anomaly detection. In *Proceedings 2001 IEEE Symposium on Security and Privacy* (pp. 130–143). IEEE. https://doi.org/10.1109/SECPRI.2001.924286

[87] Lee, W., Stolfo, S. J. & Mok, K. W. (1999). "A data mining framework for building intrusion detection models." In *Proceedings of the 1999 IEEE Symposium on Security and Privacy* (Cat. No. 99CB36344), pp. 120–132, IEEE, Oakland, CA, USA.

[88] Leland, W. E., Taqqu, M. S., Willinger, W., & Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Transactions on Networking*, 2(1), 1–15. https://doi.org/10.1109/90.282603

[89] Leung, K., & Leckie, C. (2005). Unsupervised anomaly detection in network intrusion detection using clusters. In: Proc 28th Australasian CS conf, Newcastle, Australia; 38, 333–42.

[90] Li, G., Yin, M., Jing, S., & Guo, B. (2021). An Effective Algorithm for Intrusion Detection Using Random Shapelet Forest. *Wireless Communications and Mobile Computing*, 2021, 1–9. https://doi.org/10.1155/2021/4214784

[91] Li, J., Wang, H., & Qian, Z. (2013). Behavioral traffic classification in high-speed networks: Characteristics and challenges. *Computer Networks*, 57(7), 1613–1624.

[92] Li, X., Yi, P., Wei, W., Jiang, Y. & Tian, L. (2021). "LNNLS-KH: A feature selection method for network intrusion detection." *Security and Communication Networks*, Article ID 8830431, 22.

[93] Liaw, A., & Wiener, M. (2002). Classification and regression by random Forest. *R News*, 2(3), 18-22.

[94] Liu, R., Shi, J., Chen, X., & Lu, C. (2024). *Network anomaly detection and security defense: Machine learning advances. Computers & Security, 130*, Article 103003.

[95] Liu, Y., Chen, H., & Lu, Y. (2019). A survey of anomaly detection techniques in network security. *International Journal of Computer Applications*, 19(4), 1-9.

[96] Lu, K. (2025). *Network anomaly traffic analysis.* In *Proceedings of the 1st International Conference on Cognitive & Cloud Computing (IC3Com 2024)* (pp. 37–45). Science and Technology Publications. doi:10.5220/0013344100004646.

[97] Lu, X., Liu, J., & Gan, H. (2019). *Network traffic anomaly detection combining CNN and LSTM. ACM Transactions on Internet Technology.*

[98] *Machine learning-based network anomaly detection: Point anomaly identification with SHAP interpretability.* (2024). *MDPI Journal, 5*(4).

[99] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (pp. 281-297). University of California Press.

[100] Mahoney, M. V. & Chan, P. K. (2003). "Learning rules for anomaly detection of hostile network traffic." In *Proceedings of the third IEEE International Conference on Data Mining*, pp. 601–604, IEEE, Leipzig, Germany.

[101] McCool, M., Miller, E., & Cho, J. (2014). Practical guide to NetFlow data analysis. *Journal of Computer Science & Technology,* 29(6), 811–823. https://doi.org/10.1007/s11390-014-1464-3

[102] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (pp. 1273–1282).

[103] Mohale, V. Z., & Obagbuwa, I. C. (2025). *Evaluating machine learning-based intrusion detection systems with explainable AI: Enhancing transparency and interpretability. Frontiers in Computer Science, 7*, Article 1520741. doi:10.3389/fcomp.2025.1520741

[104] Moore, A. W., &Papagiannaki, K. (2005). Toward the accurate identification of network applications. Passive and Active Network Measurement, 41–54.

[105] Mount D. (2005). KMlocal: a testbed for k-means clustering algorithms; 2005. <http://www.cs.umd.edu/~mount/Projects/-KMeans/kmlocal-doc.pdf>.

[106] Mukkamala, S., Janoski, G. & Sung, A. (2002). "Intrusion detection using neural networks and support vector machines." In *Proceedings of the 2002 International Joint Conference on Neural Networks,* 2, 1702–1707, Honolulu, Hawaii.

[107] Narmadha, S., & Balaji, N. V. (2025). *Improved network anomaly detection system using optimized autoencoder–LSTM. Expert Systems with Applications, 273*, Article 126854. doi:10.1016/j.eswa.2025.126854

[108] *Network traffic anomaly detection based on attention-driven deep DNN and OC-SVM for IoT.* (2025).

[109] Newman, M. (2010). Networks: An introduction. Oxford University Press.

[110] Nguyen, C., & Costa, A. (2025). *Anomaly detection in network traffic using machine learning techniques. ITSI Transactions on Electrical and Electronics Engineering, 13*(1), 1–6.

[111] Nguyen, T. T., & Armitage, G. (2008). A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys & Tutorials*, 10(4), 56–76.

[112] Nikolaenko, V., Ioannidis, S., Weinsberg, U., Joye, M., Boneh, D., & Taft, N. (2013). Privacy-preserving ridge regression on hundreds of millions of records. In 2013 *IEEE Symposium on Security and Privacy* (pp. 334–348). IEEE. https://doi.org/10.1109/SP.2013.28

[113] Nong, X. (2025). *Anomaly detection in imbalanced network traffic using statistical deviation techniques. Symmetry, 17*(12), 2087. doi:10.3390/sym17122087

[114] Oziegbe, T. E. (2025). *Anomaly-based intrusion detection systems in vehicle networks: Hybrid and deep learning methods. FJS Journal*.

[115] Pai, V., Pai, K., Manjunatha, S., Hirmeti, S., & Bhat, V. V. (2025). *Adaptive network anomaly detection using machine learning approaches. EURASIP Journal on Information Security, 29*, Article 29. doi:10.1186/s13635-025-00216-4

[116] Palaniappan, S., Logeswaran, R., & Khanam, S. (2025). *Anomaly detection in network traffic for insider threat identification: A comparative study. Journal of Informatics and Web Engineering, 4*(2), 145–157. doi:10.33093/jiwe.2025.4.2.10

[117] Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering,* 22(10), 1345–1359.

[118] Panda, M. Abraham, A. & Patra, M. R. (2012). "A hybrid intelligent approach for network intrusion detection." *Procedia Engineering,* 30, 1–9.

[119] Park, K., & Willinger, W. (2000). Self-similar network traffic: An overview. In *Self-Similar Network Traffic and Performance Evaluation* (pp. 1–38). Wiley.

[120] Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. *Computer Networks*, 51(12), 3448–3470. https://doi.org/10.1016/j.comnet.2006.11.001

[121] Paxson, V. (1999). Bro: A system for detecting network intruders in real-time. *Computer Networks*, 31(23-24), 2435–2463. https://doi.org/10.1016/S1389-1286(99)00115-8

[122] Portnoy, L. (2000). Intrusion Detection with Unlabeled Data Using Clustering, Ph.D. dissertation, Columbia University, New York, NY, USA.

[123] Prabowo, A. O. (2026). Evaluation of anomaly-based network intrusion detection systems with varying malicious traffic distributions. *MDPI Electronics, 6*(1), Article 14. doi:10.3390/electronics6010014

[124][125] Prasad, T. S. L., Beeram, A., & Bodige, L. (2025). Anomaly detection in network traffic using machine learning techniques. *Global Journal of Engineering Innovations & Interdisciplinary Research, 5*(4).

[125] Qu, B., Zhang, X., Li, Y., & Wang, H. (2025). *Design of network anomaly detection model based on event key time subgraphs. Symmetry, 17*(11), 1976.

[126] Rafique, S. H., Abdallah, A., Musa, N. S., & Murugan, T. (2024). *Machine learning and deep learning techniques for Internet of Things network anomaly detection—current research trends. Sensors, 24*(6), 1968. doi:10.3390/s24061968

[127] Ramadas M, Ostermann S, Tjaden B. (2003). Detecting anomalous network traffic with self-organizing maps. In: Proc recent adv intrusion detect (RAID), lecture notes in computer science, 2820, Pittsburgh, PA; 2003. 36–54.

[128] *Real-time anomaly detection in network traffic: Comparative analysis of ML algorithms.* (2025). *IRJAEH Proceedings*.

[129] Roesch, M. (1999). Snort - Lightweight Intrusion Detection for Networks. In *Proceedings of the 13th USENIX Conference on System Administration* (LISA '99) (pp. 229–238). USENIX Association.

[130] Roy, S. (2024). *A comprehensive survey on network traffic anomaly detection using deep learning.* Preprint. doi:10.13140/RG.2.2.32071.30884

[131] Santoso, N. A., Lutfayza, R., Indarmawan, B., & Gunawan, G. (2024). *Anomaly detection in network security systems using Naive Bayes and cross-validation. Journal of Intelligent Decision Support System, 7*(2), 1–13.

[132] Sattar, S., Khan, S., Khan, M. I., Akhmediyarova, A., Mamyrbayev, O., Kassymova, D., Oralbekova, D., & Alimkulova, J. (2025). *Anomaly detection in encrypted network traffic using self-supervised contrastive learning. Scientific Reports, 15*(1), Article 26585. doi:10.1038/s41598-025-08568-0

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

[134] Schummer, P., del Rio, A., Serrano, J., Jimenez, D., Sánchez, G., & Llorente, Á. (2024). *Machine learning-based network anomaly detection: Design, implementation, and evaluation. AI, 5*(4), 2967–2983. doi:10.3390/ai5040143

[135] Sculley, D. (2010). Web-scale k-means clustering. *Proceedings of the 19th International Conference on World Wide Web*.

[136] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

[137] Shiravi, H., Shiravi, A., Tavallaee, M., & Ghorbani, A. A. (2012). Toward developing a systematic approach to generate benchmark datasets for intrusion detection. *Computers & Security*, 31(3), 357–374.

[138] *siForest: Detecting network anomalies with set-structured isolation forest.* (2024). arXiv:2412.06015.

[139] Sinclair, C., Pierce, L. & Matzner, S. (1999). "An application of machine learning to network intrusion detection." In *Proceedings of the 15th Annual Computer Security Applications Conference* (ACSAC'99), pp. 371–377, IEEE, Phoenix, ARI, USA, December 1999.

[140] Snort Network Intrusion Detection System; 2006. <http://www.snort.org>.

[141] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In 2010 *IEEE Symposium on Security and Privacy* (pp. 305–316). IEEE. https://doi.org/10.1109/SP.2010.25

[142] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *Proceedings of the 2010 IEEE Symposium on Security and Privacy* (pp. 305-316). https://doi.org/10.1109/SP.2010.35

[143] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. *IEEE Symposium on Security and Privacy*, 305–316.

[144] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In *Proceedings of the 2010 IEEE Symposium on Security and Privacy* (pp. 305-316). https://doi.org/10.1109/SP.2010.35

[145] Stein, G., Chen, B., Wu, A. S. & Hua, K. A. (2005). "Decision tree classifier for network intrusion detection with GA-based feature selection*." In Proceedings of the 43rd Annual Southeast Regional Conference*, 2, 136–141, 2005.

[146] Tanenbaum, A. S., & Wetherall, D. J. (2011). Computer networks (5th ed.). Pearson Education.

[147] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications* (pp. 1–6). IEEE. https://doi.org/10.1109/CISDA.2009.5356528

[148] Tsai, C.F., Hsu, Y.F., Lin, C.Y. & Lin, W.Y. (2009). "Intrusion detection by machine learning: a review." *Expert Systems with Applications*, 36(10), 11994–12000.

[149] *Unsupervised learning algorithms for anomalous traffic detection based on flows: A systematic literature review.* (2025). arXiv:2503.08293.

[150] Verwoerd, T. & Hunt. R. (2002). Intrusion detection techniques and approaches *Comput Commun*, 25, 1356-1365.

[151] Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide (1st ed.). Springer International Publishing.

[152] Wang, J., Reddy, A. L. N., & Arif, M. (2011). Real-time anomaly detection and diagnosis for large-scale network traffic. *IEEE Transactions on Network and Service Management,* 8(4), 310–323. https://doi.org/10.1109/TNSM.2011.101511.110707

[153] Wang, P. (2025). *Abnormal traffic detection based on deep attention networks for IoT. Frontiers in Communications and Networks*.

[154] Wang, Z. Ren, X. Li, S. Wang, B. Zhang, J. and Yang, T. (2021). "A malicious URL detection model based on convolutional neural network." *Security and Communication Networks*, 2021.

[155] Yang, Z., Jin, Y., Liu, J., Xu, X., Zhang, Y., & Ji, S. (2025). *Cloud platform network traffic monitoring and anomaly detection based on large language models* (arXiv:2504.17807). arXiv.

[156] Yao, H., Liu, Y., & Chao, F. (2016). An Abnormal Network Traffic Detection Algorithm Based on Big Data Analysis. *International Journal of Computers, Communications & Control (IJCCC),* 11(4):567, 2016.

[157] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. *IEEE Access*, 5, 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418

[158] Zaharia, M., Chowdhury, M., Franklin, M. J., Shenker, S., & Stoica, I. (2016). Apache Spark: A unified engine for big data processing. *Communications of the ACM*, 59(11), 56–65.

[159] Zhang, C., Chen, Y. & Meng Y. (2021). "A novel framework design of network intrusion detection based on machine learning techniques," *Security and Communication Networks,* 2021, Article ID 6610675, 15.

[160] Zhang, J., & Zheng, Z. (2016). Anomaly detection for high-dimensional data using unsupervised learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 28(9), 2269-2281.

[161] Zhang, J., & Zulkernine, M. (2008). Anomaly based network intrusion detection with unsupervised outlier detection. In *Proceedings of the 2006 IEEE International Conference on Communications*, 5, 2388–2393. IEEE. https://doi.org/10.1109/ICC.2006.255089

[162] Zhang, J., Zulkernine, M. & Haque, A. (2008). Random forest-based network intrusion detection systems IEEE Transactions on Systems, Man, and Cybernetics – Part C. *Applications and Reviews,* 38 (5) 648-658 View in Scopus

[163] Zhang, W., & Lazaro, J. P. (2024). *A survey on network security traffic analysis and anomaly detection techniques. International Journal of Emerging Technologies and Advanced Applications, 1*(4), 8–16. https://doi.org/10.62677/IJETAA.2404117

[164]Zhang, Y., Wang, L., & Wang, X. (2017). Incremental random forests for data streams with concept drift. *In Pattern Recognition*, 69, 300–312.

[165]Zhao, X., Fok, K. W., & Thing, V. L. L. (2024). *Enhancing network intrusion detection performance using generative adversarial networks.* arXiv:2404.07464. doi:10.48550/arXiv.2404.07464.

[166]Zhou, J., Ding, L., & Li, Z. (2017). An overview of abnormal network traffic detection. *International Journal of Computer Applications*, 156(2), 1-5. https://doi.org/10.5120/ijca2017914275

[167]Zhou, M., Jin, L., & Guo, L. (2017). Real-time network traffic anomaly detection based on deep learning. Journal of Information Security and Applications, 33, 1–13. https://doi.org/10.1016/j.jisa.2017.09.002