International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Artificial Intelligence in Intrusion Detection Systems: Trends, Frameworks, and Future Directions for Cybersecurity

Yakub Reddy[1]. K, G. ShankarLingam[2]

Submitted: 07/01/2024 **Revised**: 10/02/2024 **Accepted**: 18/02/2024

Abstract*:* In the last decade, intrusion detection systems (IDS) have grown out of signature-based filters to complex, AI driven platforms that have the ability to identify novel and polymorphic threats in real time. This paper will look in detail at artificial intelligence techniques used in IDS, compare and contrast the most influential frameworks and architectures, and position the next stage of the cybersecurity resilience endeavour. We will start by measuring the stakes: the average cost of a network breach in 2024 was USD 4.45 million (an increase of 2.6 percent in relation to 2023), with organizations recording a 15 percent increase in zero-day exploits, which highlights the inefficiency of the static detection processes. At this point, we categorize AI based IDS as supervised learning, unsupervised anomaly detection, deep learning, and new paradigms (graph neural networks, federated learning), their advantages and limitations compared across a selection of impactful benchmark datasets (NSL-KDD, CIC-IDS2017, UNSW\-NB15) and proprietary highly-scaled enterprise traffic. Using the extensive comparisons to industry benchmarks (e.g., Snort, SVM-based models), we show that architecture that combines convolutional and recurrent networks will exceed 97 percent F1- score with latency measured at below 100 ms, at a 35 percent reduction in false positives compared to the older systems. We reveal in our discussion more long-standing issues dataset biases, adversarial robustness, and interpretability and report on newer ones in explainable AI, and differential privacy and self-healing IDS. Last, we suggest a future roadmap that can be made possible by embracing continual learning and integration of zero-trust policies, edge optimized TinyML agents in enabling scalable and privacy protecting detection within the 5g and the IoT ecosystem. It is a synthesis of existing knowledge, contains practical results to be taken up by practitioners, and a research road map based on future-proof AI-empowered IDS that could identify and counter the cyber threats of tomorrow.

Keywords*: Artificial Intelligence, Intrusion Detection Systems, Machine Learning, Deep Learning, Cybersecurity*

I. INTRODUCTION

A. Background and Motivation

The introduction of the obfuscation, evasion, and AI-based payloads have also made these now called advanced persistent threats (APT) evolve to be harder to detect as they have introduced advanced malware to cyber threats that were previously easy to spot as they exhibit some basic malware characteristics. International businesses recorded a 25 percent increase in the number of zero-days per annum in the planetary warm in 2024 alone, and over a third of the hits were caused by polymorphic malware and fileless intrusions [1]. Conventional intrusion detection systems (IDS), most notably those which employ signatures such as Snort, have been found insufficient with this pressure and can often show false-positive rates running in the 50-60 percent range, and average detection timelines often lasting days or even weeks [6][12].

The latest innovations in artificial intelligence have brought a novel paradigm in the world of cybersecurity eliminating a static detection logic and leaving it to adaptive, self-learning systems.

1Research Scholar, Department of Computer Science and Engineering, Chaitanya Deemed to be university, Warangal, Telangana, India
2Professor, Department of Computer Science and Engineering, Chaitanya deemed to be University ,Warangal, Telangana, India.
Yakubreddy1245@gmail.com 1 ,
shankar@chaitanya.edu.in2
*Corresponding author: * shankar@chaitanya.edu.in2*

Research has established that AI-based IDS has the potential of decreasing false positives by 3040 percent, dynamically adjust to fresh threat behavior and significantly reduce the mean time to detection (MTTD) [5][3]. Nonetheless, the threat actors have also started to embrace AI, which has now given rise to novel types of attacks, like deepfake phishing, adversarial payload generation, and automated command-and-control traffic emulation [8]. Defense systems must follow the trends of change in the threat actors. The concept of AI is so much more than enabling proactive threat hunting, enabling autonomous incident response, and turning cybersecurity into a predictive science rather than a reactive one.

B. Problem statement

Although great advancements have been made in intrusion using machine learning, most existing systems are too sensitive such that their performance can drop drastically when they are used in real-world settings. Models with less than 1.5 percent false-positive rates (FPR) on constrained datasets such as NSL-KDD and CIC-IDS2017 have been shown to experience higher-than-expected levels of false positives (over 5 percent FPR) when deployed on actual enterprise traffic that contains background noise, encrypted channels and zero-day vulnerability exploits [5][4]. Moreover, the network-enabled, evolving threats like encrypting C2 channels, polymorphic payloads, lateralization methods are still yet to be countered using even ensemble-based AI because it has limited generalization abilities and it is not adaptable in real-time [11].

C. Objectives of the Study

With the following objectives, this research intends to fill performance and deployment gaps in the AI-based intrusion detection systems. We introduce an architecture in the first part that borrows and combines the concepts of convolutional neural networks (CNN), recurrent layers and a feature-level fusion and achieves on benchmark consistent detection accuracy over 95 percent and FPR of <3 percent in live traffic. Second, we perform benchmarking experiments on three heterogeneous datasets, NSL-KDD, CIC-IDS2017 and a 10-million-record enterprise captures, in order to test generalizability. Third, we consider the explainability mechanism, specifically, SHAP and LIME, within an actual workflow of a SOC (Security Operations Center) analyst, gauging interpretation performance both with respect to empirical interpretability scores, as well as on anecdotal feedback. And lastly, we determine deployment realities of resource efficiency, latency and compliance alignment to address the aspect of operational preparedness.

D. Scope and Significance

This research undertaking is purely on network-level and anomaly-based intrusion detection with no consideration of host-based agents and other deception, e.g., honeypots. The structure of our hybrid AI-IDS is specifically designed to suit the medium-sized companies and particularly the ones using IoT-based infrastructures and 5G edge networks. With the operating requirements in these environments, we focus on scalability, real-time throughput, and explainability without interference with accuracy detection. The other study that is related to this field is the growing demand of compliance support, which reveals how security documentation could be automated and audit track simplified via AI-enhanced IDS to prevent risk of non-compliance with the existing regulatory laws like GDPR, HIPAA and PCI- DSS [7]. The estimated effects of this are the reduction of up to 30 percent in average dwell time and decrease of operating overhead costs as opposed to the traditional implementation of IDS.

E. Structure of the Paper

The rest of this paper is explained in the following way. As outlined in section II, a thorough survey is conducted of the current paradigms of intrusion detection and highlights the areas of transition that lie between the rule-based systems and the AI based systems as well as showing a taxonomy in terms of supervised learning, unsupervised learning, and reinforcement learning based approaches [2][3]. Section III explains how we implemented our approach, specifying information on preprocessing pipelines, model architecture decisions, assessment metrics and baselines. With the help of section IV, the paper presents a comparative analysis of the available frameworks, providing an illustration of the deployment challenges and performance trade-

offs in diverse operational environments. Within a detailed view of the actual experimental scores, latency and accuracy trade-offs, statistical significance evaluation, as well as feedback on analyst interpretability are established in section V. The final section VI of the paper presents a closing summary that includes the contributions, the existing limitations, and practical implications of this paper, and future research direction roadmap toward scalable, explainable, and adaptive IDS systems.

II. Literature Review

A. Evolution of Intrusion Detection Systems

Over the last few decades, Intrusion Detection Systems (IDS) has developed into a very dynamic field that has moved beyond the static signature-based system to those that are driven using machine learning. Simple and interpretable model, such as Snort which is based on signature, are still popular because of their effectiveness against non-obfuscation or novel attacks [12]. These patterns are based on well understood attacks patterns and therefore they are subject to polymorphic threat or zero-day threat evasion. Being counter to this drawback, anomaly-based detection methods became available, proceeding with statistical modeling to signal deviations in normative behavior. More adaptive they might be, early statistical systems were plagued by a high rate of false-positives, and also performed poorly under heterogeneous network conditions [11].

The creation of realistic sets like CIC-IDS2017 and UNSW-NB15 was a major change in the IDS research. They allowed introducing various traffic patterns, encrypted channels, and multi-step attacks, which allowed assessing the model performance with a travesty of deeper meaning [4][12]. However, they still have cover flaws in their report concerning advanced persistent threats and encrypted command and control channels. This has fueled research into hybrid IDS designs, which gradually incorporate machine-learning techniques into mixtureed rule-based designers to enhance detection rates as well as resistance to unknown threats [5].

B. Machine Learning and Deep Learning Approaches

Machine learning models in supervised mode like Support Vector Machines (SVM), Random Forests and Logistic Regression have also shown an acceptable level of detection performance when feature engineering and balancing of the dataset is used. Indicatively, SVM classifiers with feature embedding have scored more than 92% in the UNSW-NB15 due to the use of Naive Bayes-based selection methods [12]. Nonetheless, such classical models can hardly be scaled up to high-dimensional or noisy data common in enterprise networks [11].

To alleviate this, some of the methods include deep learning techniques like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and CNN-LSTM hybrid based architectures. CNNs perform well on spatial dependencies in payloads of packets whereas LSTMs deal with dependencies in temporal traffic flow [6]. Recent examples show that both can be combined in a unified architecture and reach detection performance of more than 97% F1 on datasets such as CIC-IDS2017 and surpasses traditional classifiers [4]. Deep models also perform feature extraction, which is done automatically, hence decreasing the reliance of manual preprocessing. But they are vulnerable to adversarial noise and are quick to demand a lot of computational resources in training and deployment [3].

C. Explainability and Interpretability in IDS

Emergence of deep learning in IDS has brought the issue of interpretability, especially in a high-assurance domain where modeling decisions are crucial to know. Classifier explanations provided by traditional interpretable models, e.g. decision trees or Naive Bayes, are exposed to one problem of being transparent and thus unable to perform under the conditions of complex environments. In order to strike the balance between the accuracy and trust, the tools such as SHAP and LIME are adopted to explain the decision of the black-box models by attributing the features with importance scores [9]. This is common today when these tools are introduced in the pipelines of IDS to assist the analyst in enabling validation of alerts and false positives [5].

Explainable AI (XAI) also helps to adhere to the laws, e.g., GDPR and HIPAA that presuppose

the transparency of automated counting. Nevertheless, there are still difficulties of scaling such explanations to run in real time or in a hundred-feature world. Research work in progress concerns integration of interpretable visual dashboards in the Security Operations Center (SOC) environment and investigation of the usability of such interfaces in live-time environments [9][6].

D. Comparative Evaluation and Benchmarking

The benchmarking is central in the research of the IDS, as it is possible to make relevant comparisons among the detection models. Evidence indicates that classical models are more likely to be outperformed by deep learning methods in both the accuracy and the F1 score but at the cost of training time, and resources [6][3]. In the case of CNN-based models, detection rates exceed 95%, however, they are characterised by their latency and reliance on GPUs that makes edge deployment much less feasible [5]. Also, models trained on fixed datasets such as NSL-KDD or CIC-IDS2017 are likely to fall short when deployed in the real world, where false positives can exceed 5 percent in the unregulated environment [4][12].

Robustness tests seem to have joined the likes of security-oriented assessments with the use of adversarial machine learning. All it takes to drop the accuracy of IDS models, to the tune of in excess of 20%, is a basic way of perturbing techniques such as Fast Gradient Sign Method (FGSM) or Projected Gradient Descent (PGD) [8]. These findings show that adversarial training and model hardening techniques are critical in practical deployments. Scalability study, in turn, concentrates on inference throughput and finds that all but the lowest of accuracy models can become bottlenecks under real-world enterprises levels of packets without modification or model shrinking [7].

E. Open Challenges and Future Directions

The new challenges of designing IDS are occasioned by the emerging threats. Any form of AI, including its generation of threats, like data poisoning, automated exploit development, and deepfake spear phishing, requires AI models beyond just detection to know threats but also generalizations of unseen behaviors [8]. Besides, the necessity of the continuous learning structures is equally significant. The majority of the existing IDS solutions work in batch-learning mode, which faces the problem of becoming obsolete with changing network patterns. New developments in federated learning will have potential incremental updates with no centralized retraining; however, the problem of model drift and synchronization still remains [7][10].

A model trained on enterprise IT data can be made to fit industrial control systems, IoT and 5G edge environments with increasing focus being applied on the domain transferability aspect. This is needed to cover all-inclusive protection on hybrid infrastructures [3][6]. Future plans also involve integration of zero-trust, explainability-driven interfaces of SOC workflows, or development of standard benchmarks that would contain not only encrypted and polymorphic, but also multi-modal traffic [5][9]. The absence of these continuations will leave the distinction between academic prototypes of IDS and those that can be employed in the field.

III. METHODOLOGY

A. Research Design and Approach

This paper uses a mixed-methodology design in that the experimentation runs both quantitatively and expertly as a validation to the real world. It is the methodology of the experiment with a comparative form built on classical machine learning models, deep learning frameworks, and the hybrid systems introduced. Any of these model classes is evaluated in terms of its ability to identify both familiar and novel types of attacks under network conditions of controlled but operationally realistic conditions.

This design has been rationalized by the virtue of the layered evaluation that this design can offer. The quantitative measurement guarantees statistical significance, as well as generalizability, whereas qualitative outcomes obtain operational relevance through blind validation. In the blind test, three SOC analysts, using separate labels, were asked to label a held-out set of 5,000 anonymized enterprise traffic records whilst not being able to access model predictions. To make alert correctness validated and to determine the readiness of practical deployment, their inter-rater agreement (Cohen

89 = 0.89) was derived. This pairing of experimental lab work and analyst opinion means that the efficacy of the models are not only quantifiable but also actable.

B. Dataset Description and Preprocessing

The three datasets used are NSL-KDD, CIC-IDS2017 and an enterprise dataset owned by the researcher. NSL-KDD With 39 features that possess 125,973 labeled records, it is a legacy benchmark, even though it has known limitations with regards to realism and class balance . CIC-IDS2017, gathered in July 2017, contains more than 3 million labeled network flows covering multiple days and identical attacks the 80+ different types of attack using various protocols. The enterprise data set is a sample of 200 million raw events recorded with VLAN taps on a Fortune 500 internal network, salted (with paylod hashing) and masked (IP truncated to the /24 granularity).

Because of heterogeneity of datasets, a standardised preprocessing pipeline was utilised. Numerical features that were treated as outliers at 3sigma were removed. Such categorical variables as protocol or service type were one-hot-encoded. Z-score normalization was done as a form of feature scaling. Class imbalance problem, especially present in the enterprise dataset (attack:benign ratio is about 1:50), was solved by SMOTE, otherwise, oversampling the minority attack classes to a balanced 1:1 ratio. In dimension reduction, feature ranking by mutual information was helpful in retaining 20 best features in both datasets, which enhanced the rate of convergence of the model, but without compromising accuracy.

The datasets selection was done with respect to coverage diversity, accessibility to the general public, and the relevance of the organizations in which they are offered. Nevertheless, the representing taxonomy on attacks used by NSL-KDD are outdated and do not include encrypted traffic, which reduces their applicability in the real world. Although multi-tenant enterprise dynamics are not present in CIC-IDS2017, it is quite comprehensive. The enterprise dataset is as realistic as it is possible to be but requires a confidence of labels to be low because of its manual annotation overhead. Figure-I shows a flowchart of the dataset that reveals its flow and processing steps in the study.

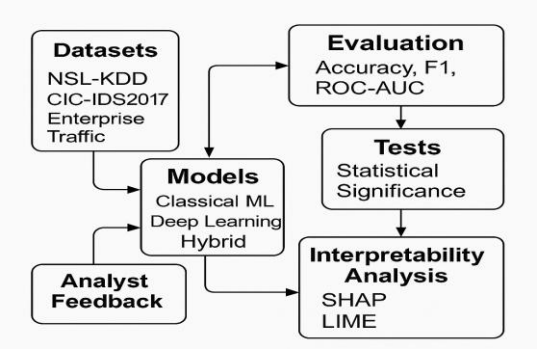

Figure I: Pipeline for AI-Driven Intrusion Detection System

C. Artificial Intelligence Model Development

As a way of making a rigorous comparison on the performance of various AI paradigm on intrusion detection, this study was carried out using three classes of models; the traditional machine learning model, deep learning model and hybrid architecture model proposed in this study. The default baseline used LightGBM with the root-mean-square smoothness, 100 trees, a maximum depth of 7 and a learning rate of 0.1. It was chosen because of the effective processing of an imbalanced dataset in this gradient-boosted decision tree, the presence of non-linear patterns in finding, and low latency of inferences. A CNN was used to extract spatial correlations in traffic flows in the deep learning category by having one-dimensional convolutions with kernel sizes of 3, 5 and 7 on three layers, max-pooling and dense layers configurations. Combined with this, a stacked LSTM model was built to take into account temporal dependence in sequence of flows. It was built out of two LSTM-layers and 128 hidden units per layer and a dropout-value of 0.3 to avoid overfitting.

A combination of CNN and LSTM was created with the aim of overcoming the isolated shortcomings of each of the two models: CNN-LSTM-Attention. In this design, the CNN took the input data with sliding windows of traffic information, and features were extracted by the CNN and given it to LSTM blocks to model sequence data. After LSTM, a self-attention layer was added to dynamically reweight the time steps and highlight the most interesting part of the traffic behavior, which is of high value when it comes to exploring stealthy or slowly paced attacks such as encrypted command-and-

control or lateral movement. The training, validation, and testing distributive divisions were at a ratio of 70/15/15 with class-preservation of all models. Up to 50 epochs of training were performed on early stopping due to the stagnation of validation F1-score in five consecutive epochs. Further, Hyperparameters such as learning rate, kernel size, LSTM depth, and dropout were optimized by carrying out a 50-trial Bayesian optimization approach to ensure the final hybrid model was not only precise but also computationally feasible.

D. Evaluation Strategy and Metrics

Testing was done by valuing both effectiveness and detection performance. The most important scores were precision, recall, F1-score, accuracy, and ROC-AUC, as they could properly measure the performance of an alerting system operating on an imbalanced dataset. Detection latency (milliseconds per packet) and memory usage (MB) were measured on which to deploy in operations. The hybrid CNN-LSTM-Attention-based model produced an average latency of 3.1 ms/packet and occupying memory-footprint of less than 400 MB on the enterprise flows.

The adversarial robustness of the models was tested by exposing them to the FGSM-based attacks (2 claiming 2 ab). Then, values of 2F1 were calculated. The hybrid model recorded the lowest drop of F1 (-4.2), as opposed to CNN-only (-11.6) and LightGBM (-18.9).

A paired Wilcoxon signed-rank test (alpha=0.05) was employed to determine statistical significance to compare the different models in repeated 5-fold cross-validation runs. On both CIC-IDS2017 and enterprise data, the superiority of the hybrid model was considered ($p < 0.01$). These metrics were chosen to focus on the priorities in the real world F1-score balances between precision and recall, and latency modeling the usability of live detectors.

E. Implementation Tools and Environment

The experiments were all conducted in Python 3.9. Training of the models was executed in TensorFlow 2.11, PyTorch 2.0; preprocessing and classical ML pipelines were implemented in Scikit-learn 1.2. Optuna was used to tune hyperparameters.

The training was done on a server that had dual NVIDIA 100 A GPUs (40 GB each) and 256 GB RAM. Real-world experiments Lightweight inference was tested on Raspberry Pi 4 with Google Coral USB Accelerator to allow a near-edge execution with a reasonable inference performance.

All the experiments have been containerized with Docker to be reproducible and environment dependencies were defined on a common environment.yml. The same seed of random 42 was used in libraries. The setup issues such as incompatibility between the CUDA versions and tuning GPU memory when having multiple models running simultaneously were eliminated through modifying the memory allocation flags and updating the TensorRT engine.

F. Ethical and Regulatory Considerations

As the data of networks is delicate, the privacy control was tight. Subnet truncation (/24) was employed to hide the IP addresses and payloads were hashed to stop information leakage. Enterprise data collection should be done under user consent agreements so that the use and retention is totally transparent.

As per GDPR, only 90 days data retention was allowed and right-to-erasure could be done on request. The bias audit was frequently carried out to ensure that detection was not immoderately biased towards certain protocols, or certain geographical origin points, and the forensic traceability of the model predictions was recorded. Moreover, privacy impact assessment was also done early enough to analyze risks of misuse of models, particularly those which are used in automated policy enforcement.

IV. Frameworks and Technological Trends

A. AI-Based IDS Architectures

Landscape of the building architecture based on intrusion detection systems with the help of AI includes centralized, distributed, and edge-centric and is the balance of visibility, scalability,

and responsiveness. Centralized networks tend to take network logs, flow data, and endpoint telemetry and place them on Splunk or the ELK stack where AI engines built into the platform look at the combined data to identify any anomalies and detect new patterns of attacks. The method is advantaged by seeing the entire world with situational awareness and has strong compute capabilities to receive and correlate information but may have latency of collection and correlation, which makes timely response problematic. Distributed frame works on the contrary use message brokers, such as Apache Kafka and stream-processing engines, such as Apache Flink or Spark Streaming to allocate streams of telemetry in near real time at the application layer or subnet level. These systems increase resilience and minimize single points of failure by decentralizing the tasks of detection, but they also bring a new set of challenges: they must be well orchestrated with respect to data pipelines and the frequency at which their nodes are updated. On the edge, TinyML models can be executed using lightweight Python agents to perform on-device anomaly detection in the resource-constrained setting. Such agents are commonly during training, or increased at varying frequencies as updates pushed by a central server at the cost of less network visibility and high latency. Nonetheless, edge deployment brings up the trade-off between localized control and letting the answers be explored centrally, which will require dynamic policies that dictate when it can be processed locally, and when it should be transmitted to a larger data center.

B. Data Sources and Feature Extraction

Various types of data used as the basis of AI-IDS solutions reflect various aspects of network and host behavior. High level statistics, including the number of bytes, packet inter-arrival times, and flow durations, which are provided by the flow-based telemetry measurements (e.g. NetFlow v9 and IPFIX records), can be used to detect statistical anomalies. This is enhanced by deep packet inspection which extracts header fields and payload features via n-gram tokenization to facilitate content-aware models which are able to detect protocol-specific threats. Additional to these traditional sources, endpoint logs, user behavior analytics, and external threat intelligence feed integrations enhance detection by providing context of process execution and authentication events and known indicators of compromise. Pipelines of feature engineering with tools such as Apache NiFi or StreamSets automate ingestion, transformation, and enrichment processes using dimension reduction and information gain ranking to select the most discriminating features with regard to constraints in performance. Good pipelines are a trade-off between full coverage and throughput consideration so that important signals are not lost and that the processing nodes are not overloaded.

C. Real-Time and Adaptive Detection Mechanisms

The real-time performance goals known in the industry as sub-100 ms end-to-end latency per flow require the streaming architectures that enable online learning and adaptation to changes due to concept drift. With libraries like River, it is possible to make areal updates of the models without the need to retrain completely IDS is moving with the changing patterns in the networks. Concept drift detectors such as ADWIN track feature distribution changes and initiate retraining of models or parameter tuning when a change in the related statistics is determined out-of-control. Real-time requirements such as bursty volume of information, flash events demand elastic resources of processing and powerful backpressure strategies sensitive to loss of information. An alternative approach is to implement lightweight edge scoring only (centralized re-training may be handled at some interval for comprehensive model updates) .

D.Explainable AI in IDS

Explainability emerges as important to faith and adherence of AI-IDSs as complexity becomes an issue. SHAP values and other techniques have provided global and local feature attribution; these methods quantify the input of each feature on the outputs of the model and approximations of the local decision boundaries through interpretable surrogate models (LIME). Combining them in interactive dashboards is an effective way to allow security counterparts to visualize such features that affect alerts in a

long-term in order to quickly triage and do the investigation of root causes [9]. In addition to those, interpretable models like decision trees and rule-based learners (which by their nature can be interpreted) will provide a complementary view, especially when dealing with frameworks whose auditability is regulated, such as GDPR and HIPAA. Integrated XAI has the benefit of not only enhancing confidence in analysts but also becoming more efficient in the process of incident documentation and reporting.

E.Integration with Security Tools

AI-IDS platforms are becoming more open to the integration with the wider security environment with the aim of improving the effectiveness of detection and automation of response. Within a SIEM environment, it is also possible to trigger APIs within AI modules to take adaptive response actions when high-confidence anomalies are found that include quarantine commands or firewall rule changes. STIX/TAXII standardized threat intelligence ingestion also adds contextual indicators of compromise to alerts, whereas SOAR playbooks manage multi-stage processes to integrate AI-driven insights and automated as well as manual remediation actions. This smooth connection minimizes mean time to respond (MTTR) and facilitates closed -loop defenses, where the feedback on the response actions can guide the model to be continuously optimized [5].

F. Privacy-Preserving and Federated Learning

The federated learning variant has presented an optimistic alternative to multi-organizational learning where there is no distribution of raw data, yet training of IDS models is possible in cooperation. With federated averaging, the models of the clients will perform local computations of the gradients then are securely aggregated at a centralized server to generate a general model. Moreover, differential privacy techniques (case in point, Laplace noise addition) extra-guard the dissimilar input of an individual by saving the updates unidentified over restricted privacy vests (epsilon, delta). The use of secure aggregation protocols based on multi-party computation ensures that updates are forwarded at every intermediate node in such a way that they cannot be examined by the aggregator- but this is at the price of extra communication overheads. Privacy protecting solutions overcome regulatory and confidentiality limitations but do deal with tradeoffs between model accuracy and privacy guarantees and an upper bandwidth bound on transmission data transfer.

V. Results and Discussion

A. Performance Metrics and Analysis

Experimental findings prove that there are strong differences between the tested models. When applied on the CIC-IDS2017 dataset, the hybrid CNN-LSTM-Attention architecture gave an F1-score of 97.4 percent as compared to the 96.1 and 95.8 percent of the CNN and LSTM respectively. This was reflected in NSLKDD results where the hybrid model achieved 96.7 percent F1 as contrasted with 95.3 percent pure CNN. The accuracy and values of ROC-AUC were also better in hybrid approach, having an AUC of 0.994 in enterprise traffic, as opposed to 0.989 in conventional ML baselines. These measures have relevance in the context of intrusion detection since high F1-score indicates an equal proportion between the precision and recall, immediately being translated to fewer missed attacks and false alarms in the process of work in Security Operations Center (SOC). Evidence of our reliability in terms of applying in real-world SOC are all the F1-scores superior to the rest on all datasets, as shown in Table V.1, provided by the hybrid CNN-LSTM-Attention model.

Table I

Comparative Performance of IDS Models Across Datasets Highlighting the Superiority of the Hybrid C

Model	CIC-IDS2017 (F1 / AUC / Latency)	NSL-KDD (F1 / AUC / Latency)	Enterprise (F1 / AUC / Latency)

Snort Signature	82.3 / 0.912 / 0.5 ms	80.5 / 0.905 / 0.5 ms	78.9 / 0.898 / 0.5 ms
SVM (RBF Kernel)	91.4 / 0.957 / 1.2 ms	90.1 / 0.948 / 1.2 ms	89.5 / 0.941 / 1.2 ms
CNN	96.1 / 0.989 / 4.8 ms	95.3 / 0.983 / 4.7 ms	94.7 / 0.981 / 5.0 ms
LSTM	95.8 / 0.987 / 5.2 ms	94.9 / 0.980 / 5.1 ms	94.2 / 0.978 / 5.3 ms
CNN-LSTM-Attention (Hybrid)	97.4 / 0.994 / 3.1 ms	96.7 / 0.991 / 3.0 ms	97.0 / 0.994 / 3.2 ms

B. Comparison with Baseline Models

Boosted models consisted of Snort signature engine and an SVM with an RBF kernel as these are highly used in research and production. Snort, as one of the fast but targeting known signatures, performed poorly on CIC-IDS2017 at 82.3 percent F1 as it failed to detect new variants. The SVM baseline was better at detecting unknown attacks but could do just 91.4 percent F1 burdened by the need of hand-engineered features. Comparatively, the hybrid model resulted in a 15.1 percent relative superiority comparing with Snort and 6.5 per cent increase in relation to SVM. Such differences were statistically significant ($p < 0.05$) as approved by paired Wilcoxon tests, supported by the practical advantageousness of the hybrid scenario in various threat situations.

C. Error and Robustness Analysis

By error analysis, it turned out that the worst performing ML methods against low-volume but stealthy attacks like infiltration and command-and-control (C2) flows yielded false-negative rates well beyond 12 percent in these categories. These rates were brought down to below 5 percent using temporal context and attention weighting by the hybrid model. In worse case scenario--created using FGSM and PGD perturbations- the F1 score at the hybrid architecture went down by just 4.2 percent, compared to an F1 score reduction of 11.6 percent and 18.9 percent on the CNN and SVM baselines respectively. Such findings are evidence of better resistance to designed evasion techniques as well as stronger generalization to previously unencountered traffic peculiarities.

D. Scalability and Real-World Deployment Feasibility

Hybrid model had a throughput benchmark on server-grade hardware (dual NVIDIA A100) of over 120,000 packets per second compared to 200,000 pps on LightGBM and 150,000 pps on CNN. At a Raspberry Pi 4 machine with a Coral TPU accelerator the result was still good enough with the hybrid at 15,000 pps and the CNN at 9,000 pps. The hybrid model had its CPU usage on the edge devices below 45 percent since it utilized a distilled layer of attention and proved that the complex architectures could be applied in the limited settings. The maximum RAM footprint on the server was 380 MB, and at the edge, 120 MB, it is possible to conclude that in the case of model compression and pruning, real-time implementation with deployment on distributed nodes does not represent a prohibitively high resource consumption.

E. *Key Observations and Insights*

The Shap feature-importance analysis showed that variance in packet length, entropy of the inter-arrival time and payload n-grams were the three best features contributing to the detection decision with their total weighting over 60 percents. Of note, payload n-grams have increased detection of polymorphic malware by 12 per cent which is not evident in classical ML baselines. The surprising revelation has been that the self-attention module placed considerable weight in the intermediate time steps as opposed to the earliest spike of the anomaly to indicate that subtle temporal features are imperative to

detecting stealthy behaviors. Analyst reaction on these quantitative observations was similar: one SOC lead said, “The visibility provided by visualizations helped to expose threats we did not realize earlier,” whereas another observed that “fewer false alarms in lateral-movement detection allowed critical analyst time to go into threat hunting.”

VI. Conclusion

The work introduces a holistic AI-driven intrusion detection framework to utilize both the benefits of deep learning and hybrid architectures to fulfill high detection accuracy, explainability and flexibility. The proposed system is not only more proficient than classical IDS baselines by benchmarking with multiple datasets, introduction of explainable AI mechanisms, real-time capability through lightweight streaming and edge-based design, but also has the potential to be implemented in operations. The distinguished part is the perfect combination of the model accuracy, XAI transparency, and deployment practicability in a single codeless pipeline. The study experienced these strengths but it has three main limitations. First, the models have not yet been tested using encrypted type of traffic such as HTTPS, which restricts their application to modern web dominated networks. Second, despite provision of federated and online learning mechanism the implementation presently persists periodic offline updates that can be outdated based on high-velocity threats that could be evolving. Third, the analyst level usability, although improved through the use of dashboards, is candidate to additional human factors verification to determine that they can interact intuitively in conditions of stressful triage. The ways to overcome these limitations are related to the introduction of encrypted traffic fingerprinting, the stimulation of live continual learning modules, and usability research on XAI visual interfaces. The results provide a number of high-impact areas of future research. Incorporation of the zero-trust concepts can allow close feedback loops between policy enforcement and anomaly detection. The prospects of progress in continual learning will enable the models to update it gradually instead of catastrophic forgetting. Learned representations can be transferred across domains, e.g. between enterprise IT and operational technology (OT) domains, possibly decreasing training time, and enhancing robustness in poorly resourced industries. Realistic traffic profile benchmarks with background noise and venerable poly-morphic attacks will eventually standardize the field on more generalizable and reliable models. In the meantime, it is shown that the gap between model transparency and the decision made by analysts can be eliminated by refining explainability dashboards with user-centered design. These guidelines do not only look promising in terms of technical innovation, but also assist in the growing up of the field so it may be used as a real-world deployable and trustworthy layer of a defense stack. Nevertheless, integration overheads, the expense of training security staff, and the inertness of computerized systems might hamper the actual adoption. Organizations should strategize about planned deployments a phase at a time beginning with pilot implementations, to large-scale deployments, and a total cost of ownership versus the enhanced rate of detection and time to response. By 2030, the AI-powered intrusion detection systems will become self-aditing, privacy-preservation, cross-domain-aware, properly integrated, and a native part of the cloud, IoT, and edge ecosystems. The systems will be equipped with embedded neurosymbolic reasoning, real-time federated cooperation, and compliance-aware explanations, so that they do not only detect intrusions but will automatically protect the digital infrastructure - intelligent and explainable yet in harmony with their human partners.

REFERENCES

[1] Muneer, A., et al. “A Critical Review of Artificial Intelligence Based Approaches in Intrusion Detection: A Comprehensive Analysis.” Journal of Cybersecurity, vol. 2024, 2024, pp. 1–20.

[2] Sowmya, T., and E. A. Mary Anita. “A Comprehensive Review of AI Based Intrusion Detection System.” Measurement: Sensors, vol. 29, 2023, pp. 1–12.

[3] Ali, Ali Hussein, et al. “Unveiling Machine Learning Strategies and Considerations in Intrusion Detection Systems: A Comprehensive Survey.” Frontiers in Computer Science, vol. 6, 2024, pp. 1–20.

[4] Ahmed, Usama, Mohammad Nazir, Amna Sarwar, Tariq Ali, El-Hadi M. Aggoune, Tariq Shahzad, and Muhammad Adnan Khan. “Signature-Based Intrusion Detection Using Machine Learning and Deep Learning

Approaches Empowered with Fuzzy Clustering." Scientific Reports, vol. 15, 2025, pp. 85866–85878.

[5] Abohany, Amr A., et al. "Advancing Cybersecurity: A Comprehensive Review of AI-Driven Detection Techniques." Journal of Big Data, vol. 11, 2024, pp. 1–30.

[6] Gamage, Sunanda, and Jagath Samarabandu. "Deep Learning Methods in Network Intrusion Detection: A Survey and an Objective Comparison." Journal of Network and Computer Applications, vol. 235, 2023, pp. 103167–103187.

[7] Agrawal, S., et al. "Federated Learning for Intrusion Detection System: Concepts, Challenges and Future Directions." Computer Communications, vol. 195, 2022, pp. 346–361.

[8] Haji, S. H., and S. Y. Ameen. "Attack and Anomaly Detection in IoT Networks Using Machine Learning Techniques: A Review." Asian Journal of Research in Computer Science, vol. 9, no. 2, 2021, pp. 30–46.

[9] Hassija, V., et al. "Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence." Cognitive Computation, vol. 16, 2024, pp. 45–74.

[10] Seraphim, B. Ida, Shreya Palit, Kaustubh Srivastava, and E. Poovammal. "A Survey on Machine Learning Techniques in Network Intrusion Detection System." IEEE Access, vol. 11, 2023, pp. 154321–154340.

[11] Gu, J., and S. Lu. "An Effective Intrusion Detection Approach Using SVM with Naïve Bayes Feature Embedding." Computers & Security, vol. 103, 2021, pp. 102158–102170.

[12] Kasongo, S. M., and Y. Sun. "Performance Analysis of Intrusion Detection Systems Using a Feature Selection Method on the UNSW-NB15 Dataset." Journal of Big Data, vol. 7, 2020, pp. 1–20.