IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Ai-Augmented Intrusion Detection Systems (Ids): Opportunities And Pitfalls

Dr.J.Jebathangam[1], Professor

Department of Computer Applications(UG), VISTAS, Chennai, Tamil Nadu, India.

Dr.R.Bhuvana[2], Professor

Department of Computer Science, Agurchand Manmull Jain College, Chennai.

Abstract: Intrusion Detection Systems (IDS) are critical components in safeguarding modern digital infrastructures against cyber threats. Initially rule-based and signature-driven, traditional IDS faced limitations in detecting novel or evolving attacks. With the integration of Artificial Intelligence (AI) and Machine Learning (ML), IDS have significantly improved in accuracy, speed, and adaptability. AI-powered IDS can detect zero-day exploits and reduce false positives through advanced anomaly detection and classification techniques. However, challenges remain—adversarial attacks can exploit ML vulnerabilities, model drift necessitates regular updates, and the use of black-box models raises explainability concerns. This paper provides a comprehensive overview of AI-augmented IDS, highlighting their advantages, common pitfalls, and potential future directions.

Keywords: Intrusion Detection System (IDS), Adversarial Attacks, Model Drift, Explainable AI (XAI), Anomaly Detection, Zero-Day Attacks,

I. Introduction

In an era where cyber threats are rapidly evolving in complexity and frequency, protecting digital assets has become a top priority for organizations. Intrusion Detection Systems (IDS) serve as a crucial line of defense by monitoring network and system activities for signs of malicious behavior. Traditional IDS, which primarily rely on predefined rules and known attack signatures, often fall short in identifying new or sophisticated threats. To address these limitations, researchers and practitioners have increasingly turned to Artificial Intelligence (AI) and Machine Learning (ML) techniques. These intelligent systems enhance IDS performance by enabling dynamic threat detection, reducing false alarms, and improving response times. This paper explores the evolution of IDS with AI integration, evaluates its benefits and shortcomings, and outlines emerging trends and future research directions in the field.

Traditional Intrusion Detection Systems (IDS) rely heavily on predefined rules or known attack signatures to identify malicious activity. While this approach offers transparency and is relatively easy to interpret, it often struggles to detect new or evolving threats, such as zero-day attacks. These systems typically require constant manual updates and expert intervention to remain effective, and they frequently generate high rates of false positives due to their rigid detection mechanisms. In contrast, AI-powered IDS leverage machine learning algorithms to analyze vast volumes of data, learn normal behavior patterns, and detect anomalies that may indicate cyber threats. These systems are more adaptive and capable of identifying previously unseen attack vectors. By reducing false alarms and improving detection speed, AI-enhanced IDS offer a more efficient and scalable solution for modern network environments. However, their use of

complex models, especially deep learning, can introduce challenges in terms of explainability and may require ongoing retraining to address model drift and maintain performance over time.

II. Tools used in IDS

Several tools have been developed to implement and support Intrusion Detection Systems, each with unique capabilities tailored for specific environments. Snort is a widely-used open-source Network-based IDS (NIDS) that utilizes a rule-driven language to detect known attack patterns in real-time. It is known for its flexibility and active community support. Suricata, another powerful NIDS, offers multi-threading, high performance, and built-in protocol analysis, making it suitable for high-speed network environments. OSSEC is a Host-based IDS (HIDS) designed to monitor and analyze activities on individual systems, including file integrity checks, log analysis, and rootkit detection. It is often used in combination with SIEM systems for centralized security management. Zeek (formerly Bro) is a powerful network analysis tool that goes beyond signature matching by providing detailed logs and contextual traffic analysis, which is particularly useful for advanced threat detection and incident response. These tools serve as foundational components in IDS implementations, enabling both traditional and AI-augmented systems to monitor, detect, and respond to a wide range of cyber threats.

III. Role of AI in IDS

Supervised Learning in IDS:

Supervised learning plays a key role in IDS by training models on labeled datasets that contain examples of both normal and malicious behavior. Algorithms such as Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks (ANN) learn to classify incoming data based on this prior knowledge. This approach is effective in detecting known attack types with high accuracy, making it suitable for environments where well-labeled training data is available.

Unsupervised Learning in IDS:

Unsupervised learning is used in IDS when labeled data is scarce or unavailable. It helps identify unusual patterns or anomalies in network behavior without prior knowledge of attack signatures. Techniques like K-means clustering and Autoencoders detect deviations from normal activity, making them valuable for identifying unknown or zero-day attacks. This approach is particularly useful in dynamic environments where threats are constantly evolving.

Deep Learning in IDS:

Deep learning enhances IDS by automatically learning complex patterns from raw data using models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). These models excel at capturing spatial and temporal relationships in network traffic or system logs, enabling accurate real-time threat detection. While powerful, deep learning models are computationally intensive and often considered black boxes due to their lack of interpretability.

Reinforcement Learning in IDS:

Reinforcement learning (RL) introduces adaptive behavior to IDS by enabling models to learn optimal responses through trial and error. In this framework, the IDS acts as an agent that interacts with its environment, receives feedback in the form of rewards or penalties, and gradually improves its decision-making. RL is especially useful in dynamic and complex network environments where predefined rules may not suffice, and real-time adaptability is crucial.

IV. Opportunities of AI-Augmented IDS

Improved Detection Accuracy

AI models can analyze complex traffic patterns and subtle anomalies that traditional rule-based IDS may miss. This leads to higher detection accuracy, especially for sophisticated or stealthy attacks.

Reduction in False Positives and Negatives

By learning from real-world network behavior, AI algorithms minimize false alarms and improve trust in alert systems. This reduces alert fatigue for security analysts and helps prioritize genuine threats.

Detection of Zero-Day and Unknown Attacks
Unlike signature-based systems that depend on known attack patterns, AI models—especially those using anomaly detection—can identify previously unseen attacks, making them more effective against zero-day exploits.

Real-Time Threat Analysis
AI-augmented IDS can process large volumes of data in real time, enabling rapid identification and mitigation of ongoing threats before significant damage occurs.

Scalability for Large and Complex Networks
AI models are capable of handling high-throughput environments such as enterprise data centers, cloud infrastructure, and IoT ecosystems. Their ability to learn from distributed data sources makes them ideal for large-scale deployments.

Automated and Intelligent Response
AI-powered systems can trigger automatic actions—such as blocking IP addresses, isolating compromised hosts, or alerting administrators—based on threat classification, enabling faster incident response.

Adaptive Learning and Self-Improvement
Machine learning models can be retrained periodically or in real-time, allowing IDS to evolve alongside new threat landscapes without manual rule updates.

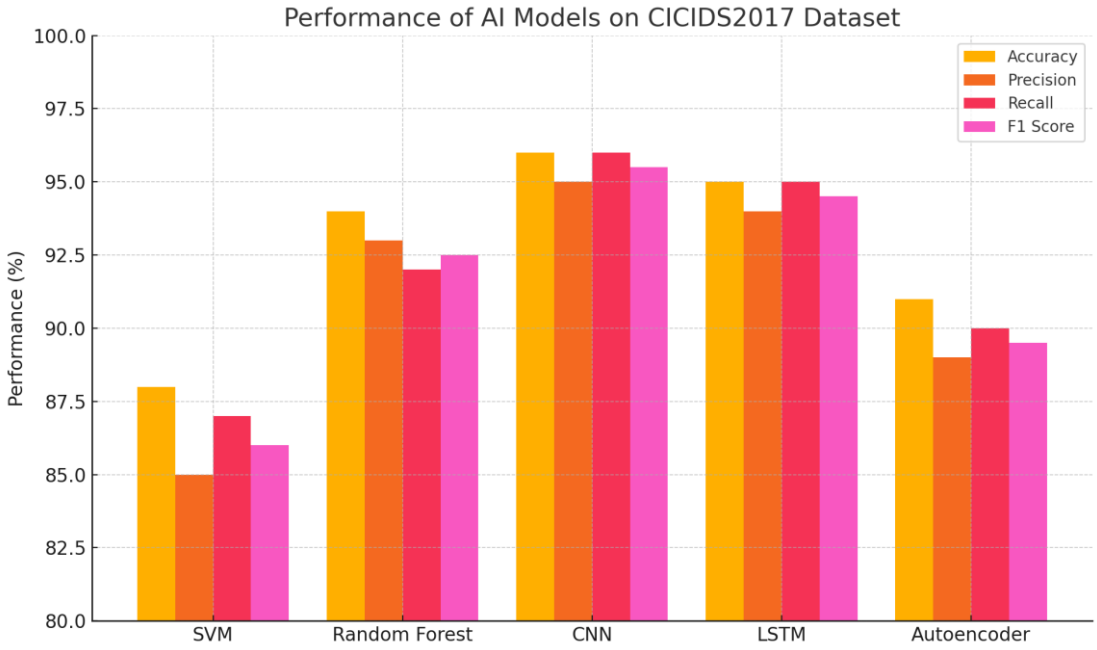
Behavior-Based Detection
AI allows systems to profile users, applications, and devices, detecting deviations from normal behavior that might indicate insider threats, account compromise, or malware activity.

V. Challenges of of AI-Augmented IDS

AI-augmented Intrusion Detection Systems, while powerful, face several significant challenges. One of the key issues is vulnerability to adversarial attacks, where malicious actors manipulate input data to evade detection. Additionally, these systems often rely on high-quality, labeled datasets, which are scarce in real-world environments. Model drift—the degradation of model accuracy over time due to changing network behavior—requires frequent retraining. Another concern is the lack of explainability, especially in deep learning models, making it difficult for security analysts to trust or interpret the results. Furthermore, deploying AI models in real-time environments can lead to performance overhead, and handling sensitive data during training may raise privacy concerns.

VI. Evaluation of AI Models

standard datasets such as NSL-KDD, CICIDS2017, UNSW-NB15, and TON_IoT. These datasets provide labeled traffic data representing normal behavior and various attack types, enabling consistent and comparative assessment across models. Key evaluation metrics include accuracy, precision, recall, F1-score, and false alarm rate. Supervised models like Random Forest and Support Vector Machines often show high accuracy on well-labeled datasets, while deep learning models such as CNNs and LSTMs excel in learning complex traffic patterns. Unsupervised models, like Autoencoders, are particularly useful for anomaly detection in unlabeled data. Realistic evaluation also considers computational efficiency, scalability, and generalization ability to unseen attack types. Cross-validation techniques and confusion matrices are commonly used to validate robustness, while testing on multiple datasets ensures model reliability in diverse network environments.



Here's a graphical representation comparing the performance of various AI models (SVM, Random Forest, CNN, LSTM, and Autoencoder) on the CICIDS2017 dataset across key metrics: Accuracy, Precision, Recall, and F1 Score. Let me know if you'd like a version for a different dataset or additional models.

VII. Conclusion

he integration of Artificial Intelligence into Intrusion Detection Systems marks a significant advancement in the field of cybersecurity. AI-augmented IDS offer numerous advantages over traditional systems, including higher detection accuracy, reduced false alarms, the ability to identify zero-day attacks, and adaptive responses in real-time. These systems leverage a variety of machine learning and deep learning techniques to analyze complex traffic patterns and evolving threats. However, the deployment of AI in IDS is not without challenges. Issues such as adversarial attacks, model drift, lack of transparency, data quality concerns, and computational overhead must be carefully addressed to ensure reliability and trustworthiness. To maximize the effectiveness of AI-augmented IDS, it is crucial to adopt strategies such as continuous learning, explainable AI, hybrid models, and secure data handling. Looking ahead, the convergence of AI with emerging technologies like blockchain and federated learning holds promise for building more robust, scalable, and privacy-preserving IDS. Thus, while AI brings transformative potential to intrusion detection, a cautious and well-informed implementation is essential for sustainable cybersecurity defense.

References

1) Mukkamala, S., Janoski, G., & Sung, A. H. (2002). Intrusion detection using neural networks and support vector machines. In Proceedings of the IEEE International Joint Conference on Neural Networks (Vol. 2, pp. 1702-1707). IEEE.
2) Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19–31.
3) Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. In ICISSP (pp. 108–116).
4) Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. In IEEE Symposium on Computational Intelligence for Security and Defense Applications.
5) Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. In IEEE Symposium on Security and Privacy (pp. 305–316).
6) Suricata IDS. (n.d.). Open Source IDS/IPS/NSM engine. Retrieved from https://suricata.io/
7) Zeek. (n.d.). Network Security Monitoring Tool. Retrieved from https://zeek.org/
8) UNSW-NB15 Dataset. (2015). Australian Centre for Cyber Security. Retrieved from https://research.unsw.edu.au/projects/unsw-nb15-dataset
9) TensorFlow. (n.d.). An end-to-end open-source machine learning platform. Retrieved from https://www.tensorflow.org/

10) Jebathangam, J., Purushothaman, S., & Rajeswari, P. (2016). Application of echo state neural network in identification of microcalcification in breast. Digital Image Processing, 8(2), 45-50.
11) PyTorch. (n.d.). Deep learning framework. Retrieved from https://pytorch.org/