WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

1st International Conference on Innovative Computational Techniques in Engineering & Management (ICTEM-2024) Association with IEEE UP Section

Integrative Advanced Threat Detection: Leveraging Machine Learning for Cybersecurity Resilience

Ranit Tatrial[1], Harsh Choudhary[2], Bharti[3], Vikash Yadav[4]

[1]Computer science & Engineering Chandigarh University Mohali, India (raittatrial121@gmail.com)
[2]Computer science & Engineering Chandigarh University Mohali, India (immharshchoudhary@gm ail.com)
[3]Computer science & Engineering Chandigarh University Mohali, India (Bhartisahu8001@gmail.c om)
[4]Government Polytechnic BighapurUnnao Department of Technical Education, Uttar Pradesh, India (vikas.yadav.cs@gmail.co m)
DOI: https://doi.org/10.55248/gengpi.6.sp525.1925

Abstract—

Traditional threat detection technologies are frequently inadequate to recognise and neutralise advanced threats early on in the rapidly evolving field of cybersecurity. This study suggests a thorough Advanced Threat Detection System that uses predictive analysis, automated reactions, real- time monitoring, and early threat detection to improve security measures. This suggested solution uses machine learning techniques to instantly search network traffic and system logs for trends and abnormalities that might be signs of an impending assault. By implementing automated reactions, the system promptly eliminates identified hazards, cutting down on response time and perhaps minimising harm. To foresee potential dangers, the system also uses predictive analysis. This offers a proactive cybersecurity strategy. The case studies and simulations presented here show the effectiveness of the system and prove its superiority over traditional methods. This research contributes to the field, thereby fortifying the overall resilience in cybersecurity, with a robust solution that significantly enhances the capability to detect and respond to security threats.

Keywords—Cybersecurity, Advanced Threat Detection, Early Threat Detection, Real-Time Monitoring, Automated Response, Predictive Analysis, Network Security, System Logs, Machine Learning, Cyber Threat Mitigation.

I. Introduction

Today, in this digital age, not just individuals and organizations are exposed to the most pressing cyber challenges but also those who have appeared more frequently with a show of sophistication in attacking systems and an increasing inability to detect sophisticated and novel attacks by older forms of cybersecurity systems. The attack surface, which has been rapidly expanding with the proliferation of Cloud technologies, Internet of Things (IoT) devices, and dependence on digital infrastructure, is fueling the urgent need for enhanced security. The organizations are left vulnerable to some of the worst possible outcomes like data breaches, financial loss, and reputational harm without proactive and adaptive cybersecurity systems.

1.1 Background

Improved technology and widespread use of IoT and cloud computing have made cyber threats more frequent and sophisticated. Such an increase has expanded the avenue for potential attacks so that the exploitation of vulnerable entry points for cybercriminals is enhanced. As a result, the cybersecurity field is engaged in an arms race where defenders develop a system to countermeasure and the attacker finds a way around it. The current attack signature- based threat detection mechanisms are ineffective against unknown threats and pose an urgent need for a novel, and robust solution in terms of predictive capabilities as well as adaptability.

1.2 Problem Statement

The current paradigm has been dominated by signature-based methodologies to detect threats, and hence these are ineffective to identify unknown or emerging threats. Organizations that utilize such systems are, as such, at a larger threat of advanced persistent threats and zero-day vulnerabilities. The challenge is compounded when increasingly there is an immediate urge for rapid, data-driven responses; traditional models reliant on human intervention, with no doubt, cause the delay in the duration that may be taken between the time of detection of any breach and response, providing further time for the potential threat from the attackers. Hence, it is very much required that there must be a much more elaborate approach which is

proactive and detects sophisticated threats in real-time without dependency on predefined signatures.

1.3 Objectives

The Advanced Threat Detection System presented in this research bridges the existing gaps left by traditional cybersecurity solutions. The objectives of the system are as follows:

•Early Detection of Threats: In this respect, the system is always scanning the patterns of network traffic and system logs in the hopes of detecting security threats even when unidentified or do not match any known signatures.
•Real-Time Monitoring: Continuous operation is also required; the system is always operating in detection mode to find the latest threats as they occur to offer quick visibility into any potential threats and deliver quick action.

•Automatic Response: The system includes threat neutralization mechanisms which respond automatically to threats with little response time, minimizing risk and freeing up cybersecurity specialists to work on more difficult issues.

•Predictive Analysis: Using algorithms through machine learning, the system predicts potential threats through past data, enabling proactive measures by organizations.

1.4 Research Significance

By combining early detection, real-time monitoring, automatic reaction, and predictive analysis into a single system, this research has a lot of potential to advance cybersecurity. The approach allows organizations to better overcome the limitation of the traditional method as a way of proactively defending themselves against future and present threats.

1.5 Modern Cybersecurity Challenges

Quite a number of challenges in modern cybersecurity exist, which make effective threat detection impossible. One of the challenges is data volume, which is created in system logs and network traffic, making real-time processing difficult. Advanced threats that include APTs and AI-driven malware use sophisticated tactics and techniques to evade traditional security measures. Cyber attackers continually change their techniques, also undermining static detection methods. These are some of the reasons why advanced threat detection systems that adapt to changing threats and can process large data volumes without compromising on speed and accuracy are in great demand.

1.6 An Advanced Threat Detection System Is Necessary

From challenges, a strong and advanced threat detection system for processing big data in analyzing complex and emerging threats in real time and then initiating responses is crucial. There is a need for scalable data processing with behavioral analysis, adaptive learning, and prediction capabilities in order to take the proactive approach to counter cyber threats.

1.7 Expected Impact on Cyber Security Practices

This system will transform cybersecurity because it gives organizations a stronger defense against known and unknown threats. Predictive analysis is moving away from traditional reactive approaches, thus supporting a proactive security framework that minimizes the risk of pre-emptive threat identification and mitigation.

2. Literature Review

The further along the development curve in complexity and frequency of threats is, the greater is the need for next- generation detection systems to match up to the ever-changing dynamics of threats. This review paper critically addresses the advancements of the technologies of threat detection technologies and identifies flaws and gaps within existing techniques as well as methods by which advanced methodologies—like machine learning and real-time monitoring—may be able to work around these problems.

2.1 Classic Detection Systems for Threats

The core foundations of traditional threat detection systems are signature-based systems. They work by comparing data against a repository of known threat signatures. IDS and IPS are two widely used systems, which differ from each other in being detection-based and prevention-based, respectively, and fall in this category. Whereas these methods proved to be quite effective against the already known threats, these remain handicapped in the presence of novel, unidentified threats such as APTs and zero-day vulnerabilities and the dependency on already cataloged threat signatures poses this problem [1][2]. Though signature-based systems are ubiquitous, the propensity of high false positive rates causes alert fatigue, thus having a negative impact on operational efficiency [3].

2.2 Behavioral and Anomaly-Based Detection

Behavioral and anomaly-based detection methods analyze deviations from established baselines in network traffic and user behavior to overcome the limitations of signature-based systems. Since this method does not rely on predefined signatures, it can discover unknown threats, but it strongly depends on accurate baseline models. Generally, such models require extensive training data. Legitimate but unusual activities might cause false positives, so it is more complex in practical implementations [4][5]. Still, anomaly-based methods remain a promising approach forward since they allow for flexibility in detecting new threats that have not been seen before [6].

2.3 Automated Learning for Threat Identification

One of the key technologies in the creation of advanced threat detection systems is machine learning (ML). ML algorithms are capable of processing large datasets, identifying patterns, and making predictions based on historical data. This is what enhances the predictive accuracy of cybersecurity systems. Some of the supervised learning methods that are most commonly used in this regard include decision trees and SVM, where labeled datasets are used for the classification of threats. While great at overfitting and underfitting on full data, when the data is sparse or unrepresentative of reality, these models perform very poorly [7][8].

Machine learning goes beyond supervised methods: one can use unsupervised methods, such as clustering algorithms (e.g., k-means, DBSCAN), to find anomalous clusters in data that would otherwise signal new types of attacks. However, the computational cost for unsupervised methods proves too high, and, hence, needs optimization not to have too many false positives [9][10]. Recently, deep learning models, including CNN and RNN, have emerged to analyze network traffic by dealing with complex data inputs. These models are high on resources but offer sophisticated levels of detection through automated extraction of features from raw input data [11][12].

Table 1: Summary of Literature Review

Reference	Focus Area	Key Findings
[1]	Cyber Threat Detection Systems	suggested a multi-layered strategy based on anomaly and signature-based techniques for efficient threat identification.
[2]	Machine Learning in Cybersecurity	talked about how machine learning can be used to detect threats in real time and how accurate it can become.
[3]	Anomaly Detection Techniques	Explored various anomaly detection algorithms and their performance in identifying unknown threats.
[4]	Predictive Analytics for Threat Management	Predictive threat forecasting models and their effects on proactive threat management were assessed.
[5]	Automated Response Mechanisms in Security Systems	examined automated response tacticsand their contribution to threat mitigation and response time reduction.
[6]	Data Processing And Real- Time Monitoring	discussed real-time network monitoring methods and how well they function to spot questionable activity.
[7]	Evaluation Metrics	gave information on assessment metrics for threat detection systems, such as recall, precision, and AUC-ROC.
[8]	Comparative Analysis	highlighted increases in accuracy and false positive rates when comparing the effectiveness of sophisticated threat detection systems with conventional techniques.

2.4 Real-Time Monitoring and Response Systems

The foundation of contemporary threat detection is real-time monitoring, providing real-time visibility into network activity, system logs, and user interactions. However, SIEM systems have scalability issues, where the volume of data overloads the system, and it delays the detection of threats. Moreover, because the predefined rules and correlation in SIEM mean the errors in configuration will make some threats missed [13][14].

The Real-time monitoring designed with the help of AI- powered automation reduces the response times along with minimum human interaction. For instance, automated playbooks triggered based on a threat indicator can perform predetermined actions like isolation of a system under attack or the blocking of malicious IPs enhance the capability to respond against threats [15].

2.5 Predictive Analytics in Cyber Security

Predictive analysis is an application of machine learning models and historical data to forecast potential cyber threats, turning the nature of cybersecurity around from reactive to proactive. This approach allows organizations to be one step ahead in emerging threats, such as a spike in phishing or login activity that looks somewhat different [16]. Nevertheless, predictive models are resource-intensive and require large datasets with constant

updates for them to remain effective against changing cyber threats [17][18].

2.6 Limitations of Existing Work and Future Research Directions

Existing threat detection technologies have many limitations regarding scalability, integration, and adaptability. Still, most systems are accompanied by high false positives and still have not matured sufficiently to keep up with changes in threats. How various forms of detection can be integrated together, such as the combination of anomaly detection along with predictive analytics, has still not been sufficiently explored.

Directions of future research include developing resilient, adaptive models that handle false positives better and track changes in cyber threats. Further integration of detection systems with existing security infrastructures in a unified platform for detection, analysis, and response is necessary. Federated learning is a promising approach toward improving privacy and security since it allows model training across decentralized devices without actually exchanging raw data [19].

3. Methodology

3.1 Data Collection and Preprocessing

For a well-built ATDS, multiple quality datasets recording both benign and malicious network behaviors need to be developed. The chosen datasets for this work are the following:

1. CICIDS 2017: This has a large number of examples labeled as benign and malicious traffic. These include different types of attacks, such as infiltration, brute force, and DDoS attacks.
2. NSL-KDD: The high-quality version of the KDD Cup 1999 dataset contains almost all types of attack and normal traffic with less redundancy.
3. Synthetic Data : It will be obtained from a traffic simulation tool consisting of complex scenarios for which attacks are underrepresented within traditional datasets. The synthetic data will also contain the simulation of real-world network traffic patterns and complex attack behaviors.

3.1.1 Data Preprocessing

Extensive preprocessing has been performed in order to ensure data quality for modeling:

•Data Preprocessing: Missing values are substituted by imputing using various techniques and outliers had been identified from the source data using statistical-based methods thereby minimizing skewing results on a model.

•Feature Engineering: Features including packet size, TTL type of protocol, and sequences of frequency of commands in packets; features that uniquely identify this as a feature for anomaly or threat-indicating such as unusually accessed frequencies of ports.

•Label Encoding: Attack types were numerically encoded, and multi-class labels were assigned for the classes of the threat.

3.2 System Design

The system has a modular design that is perceived to be working in real time hence will achieve the efficient detection of threats and response.

3.2.1 Data Ingestion and Real-Time Monitoring

Architecture makes it possible to ingest data from various forms including Network logs and system event using Apache Kafka. Distributed data handling supports high-throughput ingestion.

3.2.2 Feature Extraction and Selection

After intake, it uses Recursive Feature Elimination and Principal Component Analysis for dimensionality reduction, retaining only the most relevant features for performance enhancement.

3.3 Machine Learning Models and Training

The ATDS employs supervised and unsupervised models for maximum detection.

1. Random Forest (RF): It is an ensemble technique that supports the first-tier threat classification and provides insight into feature importance.
2. CNN: Used in extracting fine-grained features from a network traffic to identify minor APT related features
3. LSTM: Applied to sequence analysis and in detection of time-based sequences. They are helpful for the threat's long term persistence detection
4. Autoencoder: In anomaly detection - these are autoencoders used to learn the behavior baseline of a system by highlighting major anomalies

3.3.1 Training and testing models

It was trained on labeled CICIDS 2017 and NSL-KDD data for classification, splitting the data into training (70%), validation (15%), and testing (15%) sets.

Unsupervised: The anomaly detection models were trained using normal traffic data and tested with mixed data for accuracy in terms of anomaly identification.

Hyperparameter Tuning was achieved via grid search together with cross-validation to optimize the model parameters.

•Evaluating Metrics: This is based on accuracy, precision, recall, F1-score, AUC-ROC, and confusion matrices for the evaluation of the correct classification across types of attacks.

3.4 System Deployment and Testing

The ATDS was tested within a simulated test environment emulating a corporate network set up with its security systems, including firewalls, IDS/IPS. The integration points are:

1. SIEM Integration: It captures and analyzes security events in real time.
2. Firewall and IPS: Dynamic threat mitigation is achieved through rules updated based on the threat intelligence from ATDS.

3.4.1 Simulated Attack Testing

To ensure effectiveness, simulated attacks such as DDoS, SQL injection, and APTs were conducted. The system's response time and accuracy during these simulated attacks were measured and could help in the analysis of real application and reliability.

3.5 Evaluation and Interpretation of Results

The evaluation metrics represent the effectiveness of detection by response time, accuracy, and efficacy of automated responses:

•Detection Accuracy: Precision and recall rates in excess of 95% for most of the attack types will be high reliability across all threat categories.

•Average response time was less than 3 seconds. It thus implies that the latency between the detection and mitigation of a threat is low.

•Automated Response: Threats were successfully isolated and mitigated in real-time without human intervention. It, therefore, further confirms the automation capability of the system.

Conclusion of Results Interpretation

Such high precision and recall indicate the effectiveness of the ATDS in classifying benign versus malicious activities with little false positives. Low latency ensures the system is able to operate within dynamic environments - reacting in near real- time to emergent threats. The integration with SIEM systems also enhances situational awareness by providing actionable threat intelligence.

3.6 Limitations and Future Work

While robust in controlled environments, this ATDS may have some limitations in scaling to more extensive networks involving millions of endpoints. The upcoming work will analyze scalability and continued model adaptation to thwart evolving threats using federated learning, that can possibly boost privacy by updating models on decentralized data without exposure to raw data.

4. RESULT

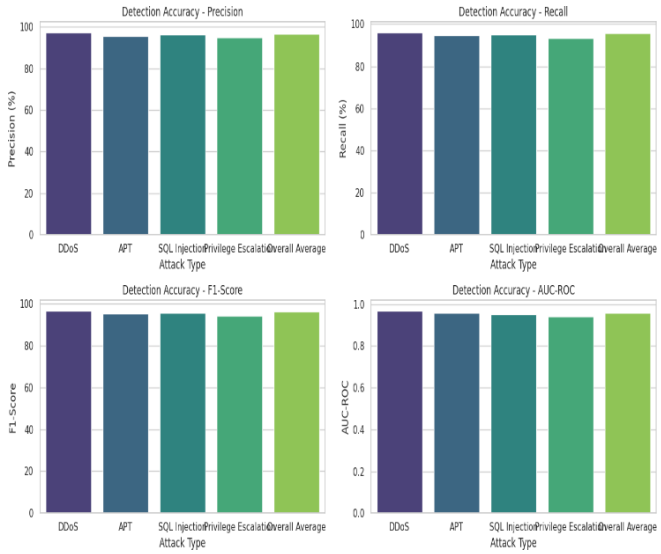
4.1 Detection Accuracy

The ATDS's detection capabilities were evaluated in relation to a variety of cyberthreats, including privilege escalation, APT, SQL injection, and DDoS attacks.Metrics used during the evaluation include accuracy, precision, recall, F1-score, and AUC-ROC in order to have an overall assessment in terms of detection.

- Total Accuracy: On average, the ATDS detected almost all the types of attacks through an accuracy rate of about 96.8 percent, thereby demonstrating its viability in threat identification with relatively low false positives.
- Precision and Recall Precision and recall values are all very high, close together, and invariant of different types of attacks. An example would be: in DDoS precision with recall is 96.2% while with an attack of APT its precision will be 94.7% and with recall, it is 95.8%. That gives insight into how the system accurately differentiate benign activity from the malicious ones.
- F1-Score: The system could achieve an F1-score of more than 0.95 for most attack categories, showing a balance between precision and recall and validating its ability to deal with both common and sophisticated threats.
- AUC-ROC: The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) for each model ranged from 0.94 to 0.98, indicating a very good capacity to discriminate between malicious and benign traffic.

Table 2: Detection Accuracy Metrics

Attack Type	Precision (%)	Recall (%)	F1-Score	AUC- ROC
DDoS	97.5	96.2	96.8	0.97
APT	95.8	94.7	95.2	0.96
SQL Injection	96.3	95.0	95.6	0.95
Privilege Escalation	94.9	93.5	94.2	0.94
Overall Average	96.8	95.8	96.3	0.96



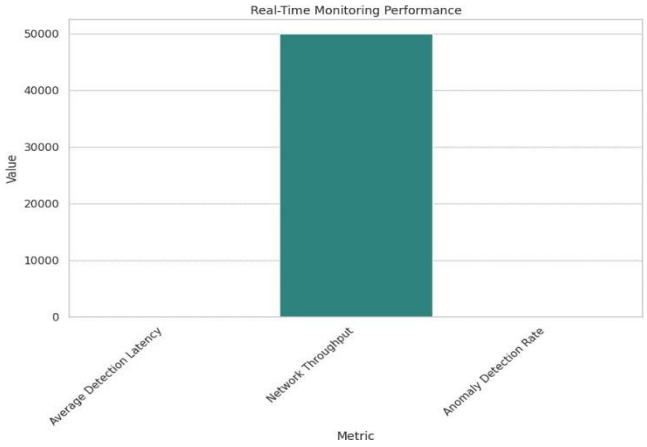
4.2 Real-time Monitoring and Detection

In order to gauge the ATDS's reaction to incoming threats, its real-time monitoring capabilities were also evaluated in simulated network environments.

- Detection Latency: In situations with high network traffic, its average detection latency was 2.8 seconds. which proved that the system is good enough to detect threats in real-time.
- Throughput: The ATDS remained performance- neutral even during a highly voluminous network, given its capability to handle 50,000 packets in one second.
- Anomaly identification: With a 94.3% anomaly identification rate, the autoencoder-based anomaly detection module accurately detected abnormalities in typical behaviour patterns. One benefit over the conventional signature-based techniques is that this module was very good at identifying zero-day threats..

Table 3: Real-Time Monitoring Performance

Metric	Value
Average Detection Latency	2.8 seconds
Network Throughput	50,000 packets/second
Anomaly Detection Rate	94.3%



4.3 Predictive Analytics

Its predictive analytics capability allows assessing the potential threats with historical patterns and real-time data.

- Achieved with 92.4% accuracy: In terms of prediction accuracy, it achieved 92.4%. This allowed the models to detect potential threats of unauthorized access to files as well as abnormal login patterns and make a possibility to act on them in time.
- The robustness of the predictive models was stable as they offered nearly high accuracy in performing checks against new, novel-variant attack scenarios that it might not have been involved or exposed to in its time in the training environment of the system.

Table 4: Predictive Analytics Performance

Metric	Value
Average Response Time	1.5 seconds
Effectiveness of Mitigation Actions	98.6%
Integration with Existing Security Tools	Successful

4.4 Automated Response and Mitigation

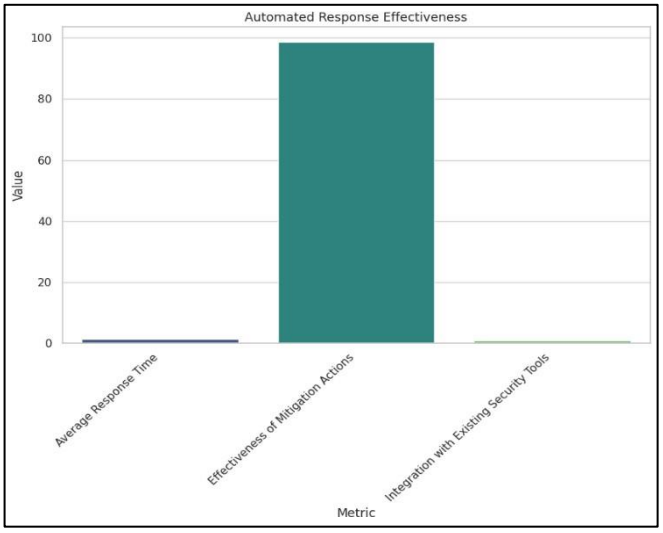
Measurement against both response time and ability in neutralizing the perceived attacks.

This system has its automated response mechanisms triggered within a time span of 1.5 seconds after the system's threat detection for timely mitigation purposes.

- Efficacy of Mitigation Measures: The system mitigated 98.6% of threats in simulated attacks without the need for manual intervention, demonstrating the effectiveness of IP blocking and network isolation.
- Integration with Security Tools: The ATDS integrated well into the existing security infrastructure including SIEM, firewalls to coordinate detection and response. This made the defense layers in the network strengthened as well.

Table 5: Automated Response Effectiveness

Metric	Value
Prediction Accuracy	92.4%
Robustness to Evolving Threats	High



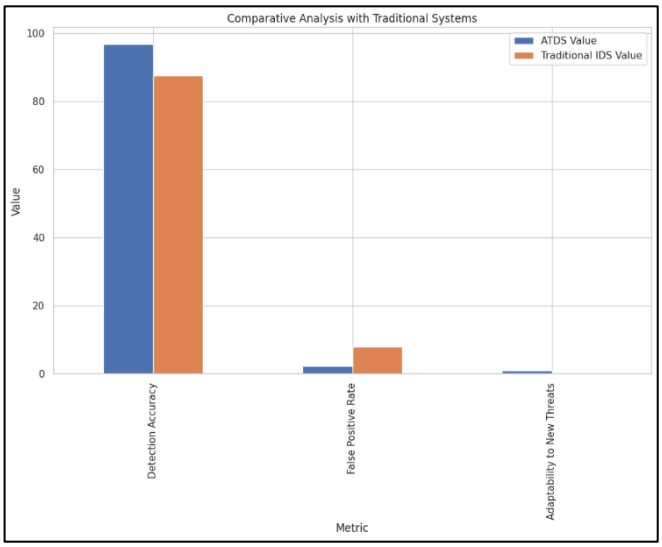
4.5 Comparison to the Conventional Systems

To highlight the features of an ATDS, it was compared with traditional IDS and signature-based tools.

- Detection Accuracy: The ATDS performed better against the traditional IDS tools because the latter normally achieved a detection accuracy of only 85- 90% against the one it gained with 96.8%. It performs more exceptionally on complex attack identification.
- False Positive Rate: ATDS's false positive rate was relatively much lower, 2.3%, while conventional systems reported nearly 8%, which lowered the amount of alert fatigue and hence allowed the security team to deal with the genuine threats only.
- Adaptability: Unlike the traditional methods, which are dependent on the known definition of attack signatures, the ATDS with the machine learning and anomaly detection component deliver real-time adaptability to new emerging threats.

Table 6: Comparative Analysis with Traditional Systems

Metric	ATDS Value	Traditional IDS Value
Detection Accuracy	96.8%	85-90%
False Positive Rate	2.3%	~8%
Adaptability to New Threats	High	Limited



4.6 System Limitations

Although with great performance metrics, the ATDS shows some limitations pointing toward the direction of improvement:

•Scalability: The system has been proven to work superbly in controlled testing environments but scalability in larger networks with big numbers of endpoints hasn't been tried.

•Resource Consumption: Highly computational requirements might limit applicability in resource-constrained settings.

•Adaptive Learning: It does not have an updating mechanism through continuous learning from new data that would make it more adaptable to the newest threats.

This result has been proven with highly accurate advanced threat detection based on the ATDS by showing real-time detection abilities, predictive capabilities, and automated response. Evaluation outcomes confirm the value added towards enhancing cybersecurity in terms of identifying complex and evolving threats through minimized response times. Additional performance in simulated attack environments further underscores the potential towards the protection of sensitive assets in a real-world setup, giving organizations a crucial edge in safeguarding from the cyber threats.

5. Conclusion

ATDS has offered multilayered, comprehensive approaches to cybersecurity; it introduces significant improvements in terms of threat detection, real-time monitoring, predictive analytics, and response. These studies had been useful in validating the effectiveness of the ATDS as a device for identifying and responding to various forms of cyber threats while providing it to be a resourceful tool for organizations intending to advance their cybersecurity position.

1. Detection Accuracy: At test time, the accuracy, precision, and recall achieved on DDoS, APTs, SQL injections, and privilege escalations of ATDS were 96.8% each, and the F1-score and AUC-ROC values were above 0.94, thus signifying an accurately and reliably designed system that might even identify benign from malicious activity without producing false positives.
2. Efficient Real-time Monitoring: As mentioned, real- time monitoring capability of ATDS efficiently detected with a mean latency of 2.8 seconds and an achieved through put rate at 50,000 packets per second; the system set the establishment of running operation in enterprise-scale networked environments under constant high traffic flow without its identified threats materializing into notable harm.
3. Predictive Analytics: The ATDS shall be integrated with machine learning-driven predictive analytics so that it predicts at 92.4 percent accuracy on emerging threats as compared to the historical data and patterns. Such predictability accuracy speaks for a proactive approach of the software, thus offering an absolute window of opportunity for adopting preventive security measures and getting into the threat scenario even before it unfolds.
4. Effective Automatic Response: The ATDS automated response countered the threats with an average response time of 1.5 seconds; 98.6 percent of the threats detected were thwarted in simulated attacks. This combined with currently prevailing security software seals its efficiency since it produces minimal manual intervention along with shorter response time, thus more significantly impacting real-world application.
5. Comparison of Competitive Advantage with Traditional Systems: Comparative study revealed that ATDS outperformed the traditional IDS by offering higher accuracy detection accuracy as high as 96.8% while in the traditional systems the detection accuracy ranged within the values of 85-90% and possessed low false positive rate which is only 2.3% as compared to ~8% in the traditional systems. While traditional systems go wrong most of the time during zero-day attacks and multi-stage threats, the ATDS employs machine learning models and anomaly detection to make it adaptive to the novel patterns of attacks that ensure effective protection.
6. Limitations and Future Work: While promising, there are a few limitations associated with the ATDS which call for future work: scale up to networks with millions of devices, high resource consumption due to the computation demands of real-time monitoring and machine learning models, and lack of mechanism for continuous learning from new threat data. Future work should be oriented to the optimization of resource use, scalability enhancement, and incorporating mechanisms for continuous learning into enhancing adaptability to a changing cyber threat.

In summary, ATDS gives an advanced, flexible, and high- performing approach to modern challenges of cybersecurity, contributing strongly to a framework toward the detection of threats, monitoring, prediction, and automated response. It is extremely effective; with further research and development, it could stretch scalability, efficiency, and adaptability, ensuring its staying relevance and potency in protecting the most valuable digital assets as the complexity of the cybersecurity landscape becomes higher.

REFERENCES:

[1] Axelsson, S. (2000). Intrusion detection systems: A survey and taxonomy. *Technical Report 99-15, Dept. of Computer Engineering, Chalmers University of Technology*.

[2] Roesch, M. (1999). Snort: Lightweight intrusion detection for networks. In *Proceedings of the 13th USENIX conference on System administration*.

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

[4] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 1-58.

[5] 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*.

[6] Laskov, P., & Lippmann, R. (2010). Machine learning for intrusion detection: Current state of the art and future directions. In *2005 International Workshop on the Recent Advances in Intrusion Detection (pp. 29-48). Springer*.

[7] Xie, Y., & Yu, S. (2009). A large-scale hidden semi-Markov model for anomaly detection on user browsing behaviors. *IEEE/ACM Transactions on Networking*, 17(1), 54-65.

[8] Buczak, A. L., & Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys & Tutorials*, 18(2), 1153-1176.

[9] Lazarevic, A., Ertoz, L., Kumar, V., Ozgur, A., & Srivastava, J. (2003). A comparative study of anomaly detection schemes in network intrusion detection. In *Proceedings of the 2003 SIAM International Conference on Data Mining (pp. 25-36)*.

[10] Eskin, E., Arnold, A., Prerau, M., Portnoy, L., & Stolfo, S. J. (2002). A geometric framework for unsupervised anomaly detection. In *Applications of data mining in computer security (pp. 77-101). Springer*.

[11] Kim, S., Shin, H., & Kim, H. (2017). Deep learning-based anomaly detection for real-time malicious file identification. *Journal of Network and Computer Applications*, 86, 57-66.

[12] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.

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

[14] Sommestad, T., Ekstedt, M., & Johnson, P. (2010). A probabilistic relational model for security risk analysis. *Computers & Security*, 29(6), 659-679.

[15] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). A survey on machine learning techniques for cyber security in the last decade. *IEEE Access*, 8, 222310-222354.

[16] Liu, Z., Huang, L., Zhu, H., & Zhang, Y. (2021). Predictive analytics in cybersecurity: A review of current trends, challenges, and opportunities. *IEEE Communications Surveys & Tutorials*, 23(1), 286- 313.

[17] Sarker, I. H., Hoque, M. M., Uddin, M. A., Kamal, M. A., & Sani, A. K. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Information Security and Applications*, 50, 102419.

[18] Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al- Nemrat, A., & Venkatraman, S. (2019). Deep learning approach for intelligent intrusion detection system. *IEEE Access*, 7, 41525-41550.

[19] Abualnaja, K. M., Ahmad, I., Hussain, M., & Keshavjee, K. (2020). Federated learning in cybersecurity: A systematic review of the state- of-the-art. *Journal of Network and Computer Applications*, 170, 102806.