Anomaly Intrusion Detection using MCA: A Review

Ms. Vatsalya[1], Prof. (Dr.) Asim Kumar[2], Prof. (Dr.) Anurag Jain
[1] Research Scholar Aryabhatta Knowledge University, Patna
[2]Director, Maulana Azad College of Engineering & Technology, Neora, Patna
[2]Director, Radharaman Engineering College, Bhopal

***Abstract*—As the complexity and size of computer networks continues to grow, securing these systems becomes increasingly difficult. Intrusion detection plays an important role in identifying and mitigating threats. This article provides a brief review of vulnerability detection techniques, focusing on their benefits, advantages, and limitations. This review covers a variety of methods for detecting anomalies, including a combination of statistical and machine learning-based approaches. Each technology is evaluated on its ability to detect previously unseen patterns and deviations from normal behavior; This makes them essential for identifying new and complex cyber threats. This article discusses the importance of feature selection and extraction in improving the performance of visual search algorithms. It explores various models of alternative attack strategies and challenges associated with mitigating vulnerability. Additionally, the review highlights the importance of real-time performance and large-scale capacity building for wide area networks. This article also discusses the role of context-sensitive information in improving the accuracy of search engine retrieval. He is exploring the integration of communication data, such as user behavior and network topology, to improve the detection capabilities of these systems. The purpose of this brief review is to provide an overview of the current state of fashion. Access to research technology, progress, trends and future research areas are discussed. By understanding the strengths and limitations of current approaches, researchers and practitioners can make informed decisions when designing efficient and effective intrusion prevention devices against changing cyber threats.**

***Index Terms*—Intrusion detection system (IDS), Signature based intrusion detection system (S-IDS), Anomaly based intrusion detection system (A-IDS),Machine Learning based detection (ML-IDS), Knowledge based detection (K-IDS), Data Mining based detection (DM-IDS),Statistical Anomaly based detection (SA-IDS), Multi classifier approaches (MCA),Adaptive and Scalable intrusion detection scheme (ASIDS).**

I. INTRODUCTION

Internet In the ever-evolving landscape of cybersecurity, the increasing sophistication of cyber threats poses a formidable challenge to the security of computer networks. As organizations and individuals become more interconnected, the need for effective intrusion detection techniques has become paramount. Anomaly-based intrusion detection has emerged as a crucial paradigm in identifying and mitigating threats that deviate from established patterns of normal behavior.

This paper aims to provide a succinct yet comprehensive review of anomaly-based intrusion detection techniques. By focusing on the effectiveness, strengths, and limitations of these methods, this review aims to contribute to a better understanding of the current state of intrusion detection in the face of evolving cyber threats.

Anomaly detection techniques play a pivotal role in identifying unknown and novel patterns of malicious activity. Unlike signature-based detection systems, which rely on a database of known attack signatures, anomaly detection methods focus on identifying deviations from normal behavior. This proactive approach is particularly crucial in safeguarding against zero-day attacks and other sophisticated intrusion attempts [1].

The paper focused to delve into various anomaly intrusion detection techniques, encompassing statistical models, machine learning approaches, and hybrid systems. We will explore the significance of feature selection, adaptation to evolving attack strategies, and the challenges associated with minimizing false positives. Additionally, the review will address the importance of real-time processing and scalability, acknowledging the complexities of securing large-scale network environments.

In summary, this brief review aims to provide insights into the current state of anomaly-based intrusion detection, fostering a deeper understanding of the challenges and opportunities in securing computer networks against emerging cyber threats.

II. INTRUSION DETECTION METHODS

Intrusion Detection Systems (IDS) are security mechanisms designed to detect and respond to unauthorized activities or potential security threats within a computer system or network. There are various methods employed by IDS to identify and analyze such activities. Here are some common intrusion detection methods [2]:

A. Signature based intrusion detection system (S-IDS)

A signature-based intrusion prevention system (IDS) is a security system that identifies known patterns or signs of malicious activity in network traffic or recorded data. These signature names are predefined and represent specific attacks or adversities. The main purpose of a signature-based IDS is to match incoming data with a signature field and trigger an alert or take action when a match is found. [2].

B. Anomaly based intrusion detection system (A-IDS)

An anomaly-based intrusion detection system (IDS) is a network security mechanism designed to identify suspicious or unusual behavior in network connections, transactions, or user behavior. Unlike signature-based IDS, which relies on attack patterns, suspicion-based IDS focuses on identifying differences in design or behavior. This approach is particularly useful in identifying previously unknown threats, including zero-day attacks and subtle, complex intrusions. [3].

C. Behavioral Analysis:

Behavioral analysis focuses on monitoring the behavior of users, applications, and systems to identify suspicious or malicious activities based on deviations from established norms[4].

D. Heuristic-Based Detection:

Definition: Heuristic-based detection involves using rules or heuristics to identify potentially malicious behavior. These rules are based on general knowledge of attack methods and techniques[4].

E. Machine Learning (ML) and Artificial Intelligence (AI):

ML and AI techniques can be applied to analyze patterns and learn from data to identify normal and malicious behaviors. These approaches can be used in both signature and anomaly detection systems[5].

F. Hybrid Approaches:

Hybrid intrusion detection systems combine multiple detection methods to improve overall accuracy and effectiveness. For example, combining signature-based and anomaly-based detection or using a combination of rule-based and machine learning techniques[5].

G. Pros and cons of anomaly detection and misuse detection are given in table 1 [6,7,8,9].

Table1: pros and cons of S-IDS and A-IDS

	Pros	Cons
S-IDS	• Accuracy in Known Threats: • Low False Positive Rates • Fast Response Time • Resource Efficiency • Prevention of Known Vulnerabilities	• Ineffectiveness Against Unknown Threats • Limited Adaptability • False Negatives • Difficulty in Encrypted Traffic Inspection: • Overhead from Regular Updates • Complexity in Rule Management • Inability to Detect Polymorphic Malware
A-IDS	• Detection of Unknown Threats • Adaptability to New Threats • Reduced Dependency on Signature Updates • Behavioral Context: • Fewer False Negatives for Unknown Threats • Customizable Baselines • Inspection of Encrypted Traffic	• Higher False Positive Rates • Complex Model Setup • Resource Intensiveness • Limited Effectiveness in Recognizing Low-Volume Attacks • Difficulty in Defining "Normal" Behavior • Inability to Detect Insider Threats

There are two main advantages of A-IDS over S-IDS. Firstly it can detect novelty i.e. “zero days” attacks. This is because of its ability to detect abnormal behaviors. Secondly A-IDS can be customized for normal activities of different systems. Thus for an attacker it becomes difficult to be aware about what kind of his activities can go on undetected [6].

III. A BRIEF REVIEW AND COMPARISON OF VARIOUS A-IDS TECHNIQUES

Anomaly-based Intrusion Detection Systems (A-IDS) aim to identify abnormal patterns or behaviors within a system or network by establishing a baseline of normal activities., the most important ones are machine learning or computational intelligence based detection (ML-IDS), knowledge based detection (K-IDS),Data Mining based detection(DM-IDS) and statistical anomaly based detection (SA-IDS). Here are various A-IDS techniques among various A-IDS techniques

1. Machine learning or computational intelligence based intrusion detection system (ML-IDS):
Intrusion detection systems (IDS) based on machine learning (ML) or artificial intelligence (CI) use advanced algorithms to identify and respond to security threats in computer networks. Here is a brief summary of the two approaches [10]:

a.Machine Learning Based IDS: Machine learning consists of training models based on historical data to make decisions or make judgments without explanation. In intrusion detection, machine learning algorithms analyze patterns in network connections or behavior to identify suspicious or known attack signatures [11].
Ambiguity detection: Machine learning algorithms are good at detecting anomalies as potential threats by understanding what is normal behavior and what is different. Techniques such as clustering, decision trees and neural networks are often used for this purpose.
Signature Detection: ML models can also be trained to identify attack domains, similar to IDS-based signatures. But machine learning expands this ability by adapting to changes in known patterns and detecting previously unseen changes [12].
Adaptability: Machine learning-based IDS is particularly effective against changing threats because the model can adapt to new attack strategies and changing network behavior. This flexibility makes machine learning ideal for detecting zero-day attacks.

b.Computational intelligence based intrusion detection system: Computational Intelligence (CI) is a broad concept that includes many software applications, including fuzzy logic, neural networks, genetic algorithms and experts. In the context of IDS, CI methods are used to deal with uncertainty, ambiguity, and decision making [13].
Fuzzy Logic Systems: Fuzzy logic allows the representation of uncertainty and ambiguity. In the context of IDS, it can be used to create rules that control the relationship between different variables and their impact on security[14].
Neural Networks: Neural networks inspired by the structure of the human brain are well-known tools. In IDS, neural networks can be trained to recognize complex patterns that indicate intrusions [15].
Genetic Algorithm: Genetic algorithm is an optimization process that results from the natural selection process. In an IDS, they can be used to optimize policies or restrictions for better access.
Expert Systems: Expert systems incorporate human experience in decision making. In the context of IDS, cybersecurity experts can contribute to decision-making and decision-making processes to ensure that systems are compatible with certain information [17].

2. Knowledge Based Intrusion Detection System (K-IDS)
Knowledge-based intrusion detection generally refers to a process that uses a knowledge base or repository of attack patterns, signatures, or behaviors to identify and segment the threat. Here is an overview of how information entry into the search function works: Knowledge Base: The system is equipped with a knowledge base containing informational articles on cognitive impairment, coping patterns, and violence-related behaviors [2].

Signature-based detection: Uses signature-based detection, where predefined signatures or known attack patterns are used to match or treat network connections.

Special detection: In addition to signature-based detection, some knowledge-based IDSs may include vulnerability detection technology. This involves establishing a baseline for normal behavior and triggering an alert or response when deviations from the baseline occur [3].

Continuous Updates: The knowledge base needs to be updated regularly to keep up with emerging threats. Security researchers and organizations are constantly looking for new attacks, vulnerabilities and malware; therefore the information needs to be updated to reflect these changes.

Alert and response: When the system detects possible interference based on patterns or anomalies, it detects and generates an alert. Depending on the configuration, it may also initiate a response such as blocking network traffic or alerting the administrator [16].

3. Data mining based approach (DM-IDS)

Data mining involves extracting patterns and information from large data sets. In intrusion detection, data mining-based methods typically include the following steps:

Data collection: Data collection documents from various sources, including network connection files, system files, and other files or relevant resources. These files often contain information about normal and potentially malicious information [17].

Preprocessing: Extract and preprocess raw data to handle missing values, normalize variables, and prepare for analysis.

Attribute Selection: Specify and select relevant properties or characteristics of the specified data. Specific selection is important to focus on the most important data and reduce the dimensionality of the data.

Data mining technology: Using data mining algorithms to analyze data first. Common data mining techniques used in migration research include Association rule mining identifying relationships and patterns in data. Clustering: Grouping similar data together [2].

Classification: Assigns samples to predefined categories (normal or negative). The goal of the system is to learn patterns and characteristics associated with these two types of activities.

Intrusion Detection: Use training models to analyze data in real time and detect suspicious or potential intrusions. Events that differ from the subject may be considered suspicious.

Creating an alarm: An alarm is generated when the system detects a suspicious or potential intrusion. These alerts can be used to alert system administrators or trigger automated responses.

Feedback Loops: Create feedback loops to extend and improve the model with new data. This helps the system adapt to changing threats and changes in the network environment [7].

4. Statistical anomaly based intrusion detection system (SA-IDS)

A Statistical Anomaly-based Intrusion Detection System (SA-IDS) is a cybersecurity tool designed to identify abnormal patterns or behaviors within a network or system. Instead of relying on predefined signatures of known threats (like signature-based IDS), SA-IDS establishes a baseline of normal behavior by analyzing historical data and network traffic. Once the baseline is established, the SA-IDS continuously monitors network activity for deviations from the norm. If it detects patterns or behaviors that significantly differ from the established baseline, it raises an alert or triggers an alarm, indicating a potential security threat or intrusion. This approach is valuable for detecting previously unknown or emerging threats that may not have known signatures. However, it requires accurate profiling of normal behavior and may generate false positives if the baseline is not properly calibrated [18].

IV. COMPARISON OF VARIOUS INTRUSION DETECTION TECHNIQUES

Comparisons are shown in table2 [2,19].

Table 2: Comparison of various intrusion detection techniques

Approach	Strengths	Weaknesses
1. SA-IDS	1. Detection of Unknown Threats 2. Adaptability 3. Behavioral Profiling 4. Minimization of False Negatives	1. False Positives 2. Baseline Calibration 3. Limited Effectiveness in High-Volume Environments 4. Difficulty Handling Seasonal or Cyclical Changes

	5. Insider Threat Detection 6. Anomaly Localization 7. Continuous Monitoring 8. Resilience to Polymorphic Threats	5. Resource Intensiveness 6. Limited Effectiveness Against Known Attacks 7. Lack of Context 8. Challenge in Identifying Slow and Stealthy Attacks 9. Difficulty Handling Encrypted Traffic
2. DM-IDS	1. Pattern Discovery 2. Behavioral Profiling 3. Flexibility 4. Scalability 5. Real-time Analysis 6. Detection of Previously Unknown Threats 7. Cross-Correlation	1. Complexity and Interpretability 2. Resource Intensiveness 3. Data Quality Requirements 4. Lack of Real-time Responsiveness 5. Adaptation to Dynamic Environments 6. Potential for False Positives 7. Data Privacy Concerns
3. K-IDS	1. Known Threat Identification 2. Low False Positive Rate 3. Faster Response to Known Attacks 4. Ease of Interpretation 5. Reduced Resource Requirements 6. Stability	1. Limited to Known Threats 2. Dependency on Regular Updates 3. Difficulty in Adapting to New Attacks 4. Susceptibility to Polymorphic Attacks 5. False Negatives for Zero-Day Attacks 6. Inability to Detect Insider Threats 7. Limited Scalability
4. ML-IDS	1. Adaptability to Changing Threats 2. Automated Pattern Recognition 3. Reduced False Positives 4. Scalability 5. Behavioral Analysis 6. Learning from Uncertainty	1. Data Requirements 2. Overfitting 3. Interpretability 4. Evading Detection 5. Resource Intensive 6. Concept Drift

V. MULTI CLASSIFIER APPROACHES (MCA)

The multi-classifier approach (MCA) involves using different types of classification or detection models to improve the overall performance of intrusion detection systems (IDS). Rather than relying on a single classifier, MCA uses diversity and optimization of multiple classifiers to improve accuracy, robustness, and flexibility. Each classifier will focus on different aspects of intrusion detection, helping to provide greater protection against different types of attacks. Multi-classifier approaches involve combining the outputs of multiple classifiers to improve overall system performance in tasks such as classification or decision-making. These approaches are often used to enhance accuracy, robustness, and generalization of a system. Here are some common multi-classifier techniques:[20]

Ensemble Learning: Ensemble methods combine predictions from multiple base classifiers to form a more accurate and robust model. Popular ensemble techniques include Bagging (Bootstrap Aggregating) and Boosting.

Voting Systems: Different classifiers "vote" on the class label of a given instance, and the class with the majority of votes is assigned. This can be done through techniques such as majority voting, weighted voting, or soft voting (weighted by confidence scores).

Stacking (Meta-Ensemble): Stacking involves training multiple base classifiers, and a meta-classifier is trained to combine their outputs. The meta-classifier learns how to best weigh the predictions of the base classifiers.

Boosting: Boosting focuses on training multiple weak learners sequentially, with each new learner correcting the errors of the previous ones. AdaBoost and Gradient Boosting are common boosting algorithms.[2]

Random Forest: Random Forest is an ensemble method that constructs a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [21].

Cascade Classifiers: Cascade classifiers consist of a sequence of classifiers where each classifier in the cascade is trained to reject non-target instances quickly. If a classifier passes an instance, it is then passed to the next classifier in the sequence.

Error-Correcting Output Codes (ECOC): ECOC represents multi-class classification problems as a series of binary classification problems. Each class is assigned a unique binary code, and multiple binary classifiers are trained, each specializing in distinguishing between classes based on their codes.
Hybrid Approaches: Hybrid methods combine classifiers from different paradigms (e.g., combining decision trees with neural networks) to leverage the strengths of each component

	Pros	Cons
Ensemble Learning	Improved accuracy, robustness, and generalization.	Increased computational complexity and potential overfitting if not carefully tuned.
Voting Systems	Simple to implement, effective in improving accuracy.	May not handle well-calibrated confidence scores, and the choice of voting scheme is crucial.
Stacking (Meta-Ensemble)	Can capture diverse patterns in the data, potentially outperforming individual classifiers.	Requires additional training steps, may be computationally expensive.
Boosting	Effective at reducing bias and improving accuracy.	Sensitive to noise and outliers, potential overfitting.
Random Forest	Robust against overfitting, handles high-dimensional data well.	Can be computationally expensive during training.
Cascade Classifiers	Efficient for tasks where negatives significantly outnumber positives.	Sensitivity to the order of classifiers in the cascade.
Error-Correcting Output Codes (ECOC)	Handles imbalanced class distributions, accommodates non-linear decision boundaries.	Increased complexity and potential challenges in combining outputs effectively.
Hybrid Approaches	Can exploit complementary strengths of different classifiers.	Increased complexity and potential challenges in combining outputs effectively.

Benefits of Multi Classifier Approaches:
Improved Accuracy: Combining the outputs of multiple classifiers can lead to more accurate and reliable intrusion detection by leveraging the strengths of each classifier [22].
Robustness: MCA is often more robust against variations in data and attack strategies, as different classifiers may excel in different scenarios.
Adaptability: The adaptability of MCA allows it to handle changes in the network environment, emerging threats, and evolving attack patterns more effectively.
Reduced False Positives/Negatives: By combining diverse classifiers, MCA can mitigate the weaknesses of individual classifiers, leading to a reduction in false positives and false negatives.

Challenges of Multi Classifier Approaches:
Complexity: Managing multiple classifiers and their interactions can introduce complexity in terms of training, integration, and system maintenance.
Computational Overhead: MCA may require additional computational resources, especially when training and combining multiple classifiers [23].
Selection of Classifiers: Choosing appropriate classifiers and ensuring their diversity is a critical challenge in designing an effective MCA system.

Intrusion detection systems employing Multi Classifier Approaches aim to enhance the overall performance and reliability of the system by leveraging the collective intelligence of multiple classifiers.

VI. CONCLUSION

In conclusion, employing multi-classifier approaches in anomaly intrusion detection provides a robust and adaptable solution to the complex challenges posed by cybersecurity threats. By combining the strengths of different classifiers, these approaches enhance detection accuracy, generalization, and resilience, making them valuable tools for securing modern IT environments. However, it's essential to carefully design and configure multi-classifier systems, considering factors such as classifier diversity, data distribution, and the specific requirements of the intrusion detection task. Regular updates and retraining of the classifiers are also crucial to ensuring continued effectiveness against emerging threats.

REFERENCES

[1] Thudumu, S., Branch, P., Jin, J. et al. A comprehensive survey of anomaly detection techniques for high dimensional big data. J Big Data 7, 42 (2020). https://doi.org/10.1186/s40537-020-00320-x

[2] Anurag Jain, Bhupendra Verma, and J. L. Rana. "Anomaly intrusion detection techniques: a brief review." International Journal of Scientific & Engineering Research 5.7 (2014): 1372-1383.

[3] Xu, Xiaodan, Huawen Liu, and Minghai Yao. "Recent progress of anomaly detection." Complexity 2019 (2019).

[4] Mohindru, Vandana, and Shafali Singla. "A review of anomaly detection techniques using computer vision." Recent Innovations in Computing: Proceedings of ICRIC 2020 (2021): 669-677.

[5] Fernandes, Gilberto, et al. "A comprehensive survey on network anomaly detection." Telecommunication Systems 70 (2019): 447-489.

[6] Basora, Luis, Xavier Olive, and Thomas Dubot. "Recent advances in anomaly detection methods applied to aviation." Aerospace 6.11 (2019): 117.

[7] Abusitta, Adel, Martine Bellaiche, Michel Dagenais, and Talal Halabi. "A deep learning approach for proactive multi-cloud cooperative intrusion detection system." Future Generation Computer Systems 98 (2019): 308-318.

[8] Bhatnagar, Amrita, Arun Giri, and Aditi Sharma. "A Hybrid Intrusion Detection Approach for Cyber Attacks." Journal of Cybersecurity & Information Management 13, no. 2 (2024).

[9] Lundin, Emilie, and Erland Jonsson. "Anomaly-based intrusion detection: privacy concerns and other problems." Computer networks 34, no. 4 (2000): 623-640.

[10] Al-Ambusaidi, Mohammed, Zhang Yinjun, Yar Muhammad, and Abid Yahya. "ML-IDS: an efficient ML-enabled intrusion detection system for securing IoT networks and applications." Soft Computing 28, no. 2 (2024): 1765-1784.

[11] Halimaa, Anish, and K. Sundarakantham. "Machine learning based intrusion detection system." In 2019 3rd International conference on trends in electronics and informatics (ICOEI), pp. 916-920. IEEE, 2019.

[12] Abubakar, Atiku, and Bernardi Pranggono. "Machine learning based intrusion detection system for software defined networks." In 2017 seventh international conference on emerging security technologies (EST), pp. 138-143. IEEE, 2017.

[13] Gupta, Abhishek, Om Jee Pandey, Mahendra Shukla, Anjali Dadhich, Samar Mathur, and Anup Ingle. "Computational intelligence based intrusion detection systems for wireless communication and pervasive computing networks." In 2013 IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-7. IEEE, 2013.

[14] Shanmugavadivu, R., and N. Nagarajan. "Network intrusion detection system using fuzzy logic." Indian Journal of Computer Science and Engineering (IJCSE) 2, no. 1 (2011): 101-111.

[15] Jirapummin, Chaivat, Naruemon Wattanapongsakorn, and Prasert Kanthamanon. "Hybrid neural networks for intrusion detection system." In Proc. of ITC–CSCC, vol. 7, pp. 928-931. 2002.

[16] Hoque, Mohammad Sazzadul, Md Abdul Mukit, and Md Abu Naser Bikas. "An implementation of intrusion detection system using genetic algorithm." arXiv preprint arXiv:1204.1336 (2012).
[17] Añasco, Cesar, Karen Morocho, and María Hallo. "Using Data Mining Techniques for the Detection of SQL Injection Attacks on Database Systems." Revista Politécnica 51, no. 2 (2023): 19-28.
[18] Bindra, Shivam, and Aruna Malik. "An analysis of anomaly detection techniques for IoT devices: A review." In 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), pp. 275-280. IEEE, 2023.
[19] Chauhan, Himadri, Vipin Kumar, Sumit Pundir, and Emmanuel S. Pilli. "A comparative study of classification techniques for intrusion detection." In 2013 International Symposium on Computational and Business Intelligence, pp. 40-43. IEEE, 2013.
[20] Larriva-Novo, Xavier, Carmen Sánchez-Zas, Víctor A. Villagrá, Mario Vega-Barbas, and Diego Rivera. "An approach for the application of a dynamic multi-class classifier for network intrusion detection systems." Electronics 9, no. 11 (2020): 1759.
[21] Corona, Igino, Giorgio Giacinto, and Fabio Roli. "Intrusion detection in computer systems using multiple classifier systems." Supervised and Unsupervised Ensemble Methods and their Applications (2008): 91-113.
[22] Trapeznikov, Kirill, Venkatesh Saligrama, and David Castañón. "Multi-stage classifier design." In Asian conference on machine learning, pp. 459-474. PMLR, 2012.
[23] Benediktsson, Jon Atli, Jocelyn Chanussot, and Mathieu Fauvel. "Multiple classifier systems in remote sensing: from basics to recent developments." In International workshop on multiple classifier systems, pp. 501-512. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.