Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning And Deep Learning

Ms.M.Kowsalya[1], Ms.A.Meena[2], Ms.D.Mohanapriya[3], Ms.C.Sowmiya[4]
[1] Assistant Professor, Department of Computer Science and Engineerin,
[2, 3, 4] Final Year, Department of Computer Science and Engineering
[1,2,3, 4] Erode Sengunthar Engineering College (Autonomous) Thudupathi, Erode, Tamil Nadu, India

Abstract- *The dynamic issues in cyber security are examined via the lens of intrusion detection, using the ADT-SVM (Adaptive Decision Tree-Support Vector Machine) method. In the context of a fast evolving cyber threat scenario assisted by the Internet, the research investigates the use of Machine Learning (ML) approaches, highlighting the importance of data. The researchers index, study, and analyse publications presenting various ML algorithms, with an emphasis on temporal or thermal correlations, while also highlighting widely used network datasets and the issues connected with ML in cybersecurity. Using the KDD dataset as a benchmark, the project uses the ADT-SVM method to divide data properties into four categories: Basic, Content, Traffic, and Host. Evaluation measures, such as Detection Rate (DR) and False Alarm Rate (FAR), are then used to evaluate the performance of an Intrusion Detection System.*

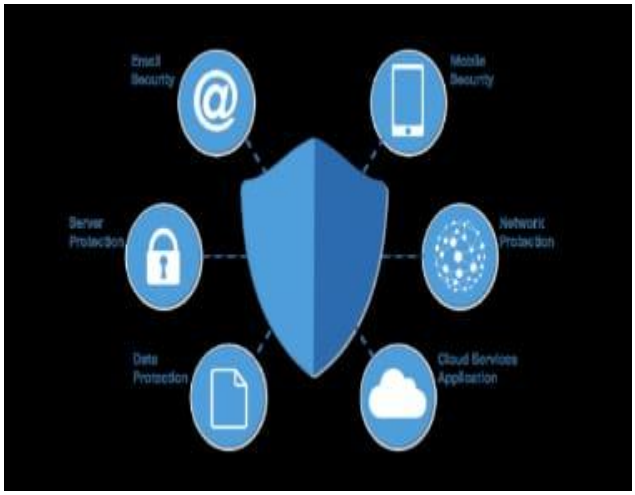
Keywords- Network Security, Cyber Threats, Anomalies, Machine Learning.

I. INTRODUCTION

The mitigation and identification of cyber threats have become critical in the ever changing field of network security. Innovative strategies are often required since traditional tactics are unable to keep up with the sophistication of contemporary attackers. Because it can identify patterns and abnormalities in large datasets, machine learning (ML) has become a powerful tool for strengthening network defences. In order to improve the proactive detection of possible attacks, this introduction investigates the integration of machine learning techniques in the context of network security. Through the utilization of sophisticated algorithms, machine learning (ML) presents the prospect of more flexible, effective, and expandable solutions, ushering in a new phase of increased cyber-attack resistance.

1.1 NETWORK SECURITY

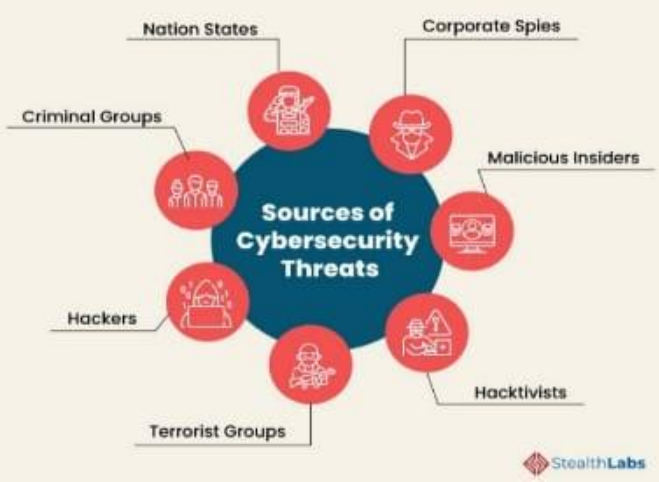
Network security is an essential component of modern digital environments, protecting the availability, confidentiality, and integrity of data transferred between linked systems. Given the widespread reliance on interconnected networks in today's world, it is more important than ever to protect these infrastructures from a wide range of potential dangers. Network security is a complex field that includes the use of intrusion detection systems, encryption protocols, firewalls, and other strong defences. Malicious actors' techniques also evolve with technology, therefore network security strategies must stay innovative and constantly changing. This introduction explores the core significance of network security and clarifies its function as the first line of defence against a wide range of cyber threats that aim to take advantage of weaknesses in the complex web of interconnected digital ecosystems.



1.2 CYBER THREATS

Cyber dangers are a real concern to the integrity and security of information systems in our digitally connected and interconnected society. Cyber threats comprise a broad range of malevolent actions planned by individuals, collectives, or states with the aim of jeopardizing privacy, causing disruptions, and taking advantage of weaknesses in computer networks. Cyber dangers are constantly changing in terms of sophistication and scope, ranging from advanced hacking methods to social engineering strategies. Our digital infrastructure is interconnected, which increases the potential

effect of these risks and makes them a widespread worry for governments, businesses, and individuals. This introduction examines the complexity of cyber threats and highlights the need for all-encompassing cybersecurity solutions to reduce risks and protect the integrity of our data-driven and more interconnected society.



1.3 ANOMALIES

Anomalies are variations or abnormalities from the expected or usual patterns in a variety of disciplines, including data analysis, system monitoring, and network security. These variations may point to underlying problems, possible hazards, or areas in which more research may be needed. Anomalies can indicate anything from mistakes in data collection to new and previously unknown patterns, making them important signals that require attention. Finding and interpreting abnormalities is critical in a variety of domains, from spotting possible security breaches in network data to detecting irregularities in financial transactions. The relevance of anomalies as departures from the norm is examined in this introduction, with a focus on how they can reveal hidden patterns, possible dangers, and areas that need more research in a variety of analytical and surveillance domains.

Objective

- Detect network intrusions with high accuracy. This means that the IDS system should be able to identify both known and unknown attacks with a low false positive rate.
- Reduce over fitting. Over fitting occurs when the IDS system learns the training data too well, which can lead to poor performance on new data. The MRF algorithm uses bagging and random feature selection to reduce over fitting.
- Improve adaptability and flexibility. The proposed system automatically selects the studied parameter values according to the used training dataset, which makes the system more adaptable to different types of network traffic and intrusion patterns**.**

II. RELATED STUDY

In this study, Abebe Diro [1] et al. have proposed The vast array of smart gadgets that make up the Internet of Things (IoT) are able to gather, store, process, and communicate data. The Internet of Things' adoption has created a wealth of potential for innovation in businesses, homes, the environment, and industries. However, worries about widespread adoption and applications have been raised by the IoT's inherent weaknesses. In contrast to conventional IT systems, the Internet of Things (IoT) ecosystem presents security challenges because of the distributed nature of smart devices, resource limitations, and heterogeneity. As a result, host-based preventive techniques like antivirus and anti-malware software cannot be used. Due to these difficulties and the nature of Internet of Things applications, monitoring systems like anomaly detection are required at the device and network levels, extending beyond organizational boundaries. This implies that, compared to other security measures, anomaly detection systems are in a strong position to secure Internet of Things devices. Our goal in this study is to present a thorough analysis of previous efforts in creating machine learning-based anomaly detection systems for IoT system security. Furthermore, we show that blockchain-based anomaly detection systems are able to jointly develop efficient machine learning models for anomaly detection. [1]
Kewen Li [2] et al. have proposed A popular ensemble learning framework, the Adaptive Boosting (Ada Boost) algorithm produces strong classification results on a variety of datasets. Unfortunately, because the Ada Boost technique is primarily meant to process misclassified samples rather than samples of minority classes, it can be difficult to apply it directly to imbalanced data. This paper proposes an improved Ada Boost algorithm (Ada Boost-A) based on AUC, which improves the error calculation performance of the Ada Boost algorithm by comprehensively considering the effects of misclassification probability and AUC. The purpose of introducing the indicator Area Under Curve (AUC) is to better process imbalanced data. In order to mitigate the generation of redundant or useless weak classifiers by the traditional Ada Boost algorithm, this paper presents PSOPD-Ada Boost-A, an ensemble algorithm that can optimize the coefficients of Ada Boost weak classifiers and re-initialize parameters to prevent falling into local optimum. [2].
Mengyao Zhu [3] et al. have proposed These days, ensemble learning is a widely used technique in machine learning-

based intrusion detection systems to increase detection accuracy. Sadly, the accumulation and reuse of past information, as well as the susceptibility of the detection model to various forms of attacks, have not been taken into account in the studies that have already been done, which results in low detection accuracy. This research suggests a model based on sustainable ensemble learning to address the problem. During the model training phase, we construct multi-class regression models that enable ensemble learning to adapt to various threats by using the probability output and classification confidence of each individual classifier as the training data. Additionally, an iterative updating strategy is described for the updating step, wherein the parameters and decision outcomes of the historical model are included into the new ensemble model's training process to achieve incremental learning. The findings of the experiment demonstrate that the suggested model performs noticeably better in terms of detection accuracy, false alarm, stability, and robustness than the current solutions. As network-based computer services and applications progress, there is a growing number of security threats on the Internet.[3].

Jinjie Liu [4] et al. have proposed Significant increases in wireless network traffic result from the integration of mobile technologies and Internet of Things (IoT) enabled devices into our daily lives, which has led to the generation of a vast scale of high dimensional network log data. This has created difficulties for Wi-Fi network security systems, which must now evaluate extremely complicated huge data in order to detect intrusions. Intrusion Detection Systems (IDS) with machine learning capabilities are a frequent feature of many Wi-Fi network systems. These IDS typically use supervised techniques that primarily rely on human experts' observations throughout the training data classification process' labelling, feature extraction, and feature selection stages. In this study, we propose an unsupervised approach with automatic feature extraction and selection process to replace human intervention and manual labelling process for analysing a large scale high dimensional data to improve the prediction accuracy of classification to detect the three most common types of network attacks: injection, flooding, and impersonate attacks in an IDS with a large scale of high dimensional data. This is done using the recently collected Aegean Wi-Fi Intrusion Dataset (AWID), which contains real traces of different types of network attacks. [4].

Leila Mohammad pour [5] et al. have proposed Internet applications have developed and become increasingly popular during the last few years. This has made the need for secure Internet networks even more imperative. Network security depends on intrusion detection systems (IDSs), which use artificial intelligence (AI) techniques. Deep learning (DL) techniques are being successfully used in IDSs as a subfield of AI. Convolutional neural networks (CNNs) are a popular type of deep learning neural network structure that are used to handle complex data. CNN is widely utilized in intrusion detection systems (IDSs) and circumvents the usual drawbacks of traditional machine learning techniques. IDSs use a variety of CNN-based techniques to address security risks and privacy concerns. Nevertheless, to the best of our knowledge, no thorough surveys of IDS programs have made use of CNN. In order to improve our comprehension of the several applications of the CNN in identifying network intrusions, abnormalities, and other forms of attacks, the main focus of this study is on CNN-based intrusion detection systems. This work summarizes the main features and contributions of the examined CNN-IDS techniques by creatively organizing them into different categories. [5].

Felix Obiteet.al. has presented in this research paper that the significant growth in Internet traffic confirms the shift of the telecommunications backbone from time division multiplexing (TDM) to a focus on Ethernet solutions. Ethernet PON, which combines low-cost Ethernet and fiber infrastructures, has emerged as the dominant technology in a market previously dominated by DSL and cable modems. This new technology is characterized by its simplicity, affordability, and scalability, enabling the delivery of massive data services to end-users over a single network. The paper provides an overview of the evolution of Ethernet Passive Optical Network (EPON), with a particular emphasis on the ongoing development of future high-data-rate access networks such as Next-Generation Passive Optical Network Stage 2 (NG-PON2), Wavelength Division Multiplexing (WDM) PON, and Orthogonal Frequency Division Multiplexing (OFDM) PON. Additionally, the recently concluded 100 Gb Ethernet Passive Optical Network (100G-EPON) is reviewed to highlight the latest advancements in the field. This comprehensive and up-to-date review aims to equip network operators and interested practitioners with a clear understanding of common priorities and timelines. Furthermore, the study aims to identify technical solutions for future investigation. The exponential increase in data traffic and the growing number of online users, who spend more time online and engage in bandwidth-intensive applications, necessitate broadband services that can support high-speed internet transmission. Therefore, future access networks must possess large bandwidth capacity and mobility to accommodate new and real-time broadband applications. DSL and cable modems are inadequate to meet such demands.[6].

Recently, Somayye Hajiheidariet.al. proposed a system that introduces a new aspect of intelligent objects by reducing the power consumption of electrical appliances. This advancement allows everyday physical objects to be enhanced with electronic devices, enabling them to connect to the internet and possess local intelligence. This concept is referred to as the Internet of Things (IoT), which encompasses these intelligent objects. However, due to their direct connection to the internet, these objects are susceptible to attacks from malicious individuals. The accessibility of resource-constrained devices through public internet access exposes them to potential intrusions. These intrusions, known as internal attacks, do not explicitly damage the network but infect internal nodes to carry out attacks on the network. Therefore, the implementation of Intrusion Detection Systems (IDSs) in the IoT is crucial. Despite the significance of this topic, there is currently no comprehensive and systematic review that discusses and analyzes the mechanisms of IDSs in the IoT environment. Hence, this paper presents a Systematic Literature Review (SLR) of IDSs in the IoT environment.[8].

BayuAdhi Tamaet.al. has proposed a system that emphasizes the importance of Intrusion Detection Systems (IDSs) in preventing cyberattacks. In order to enhance the detection rate, there is a need to develop an improved detection framework, especially when utilizing ensemble learners. The process of designing an ensemble faces two main challenges: selecting appropriate base classifiers and combiner methods. This research paper provides an overview of how ensemble learners are utilized in IDSs through a systematic mapping study. We have gathered and analyzed 124 significant publications from the existing literature. These publications have been categorized based on the year of publication, publication venues, datasets used, ensemble methods, and IDS techniques. Additionally, this study presents an empirical investigation of a novel classifier ensemble approach called "stack of ensemble" (SoE) for anomaly-based IDS. The SoE is an ensemble classifier that employs a parallel architecture to combine three individual ensemble learners, namely random forest, gradient boosting machine, and extreme gradient boosting machine, in a homogeneous manner. The performance of different classification algorithms is statistically evaluated using metrics such as Matthews correlation coefficients, accuracies, false positive rates, and area under the ROC curve.[9].

Muhamad Erza Amina et al. have presented a system that addresses the security challenges posed by the widespread use of IoT-enabled devices in our daily lives, thanks to recent advancements in mobile technologies. The main concern lies in the vulnerability of wireless mediums like Wi-Fi networks, which are open in nature. An impersonation attack occurs when an adversary disguises themselves as a legitimate party within a system or communication protocol. The abundance of connected devices generates a vast amount of high-dimensional data, making simultaneous detections complex. However, feature learning can mitigate potential issues arising from the large volume of network data. In this study, a novel approach called Deep-Feature Extraction and Selection (D-FES) is proposed, which combines stacked feature extraction and weighted feature selection. By reconstructing relevant information from raw inputs, stacked autoencoding enhances the meaningfulness of representations.[10].

III. EXISTING SYSTEM

It was chosen to provide an interconnection strategy that permits information sharing and communication relationships without the need for human intervention. The Internet of Things architecture has made it possible for different devices to be connected for a crucial amount of time without the need for human intervention. There are less pre-arranged ideas to obtain information, and the amount of information has decreased, which was not the case before. As an illustration, consider enhanced attack and various edges. However, a few theories, such as artificial mindfulness, artificial intelligence, and significant learning, have a lot to say about their potential as well as the verified benefits of preparing heterogeneous information of different estimations and different specialists expected to treat it. Based on the findings, the proposed research project has employed significant learning hypothesis to select a security understanding for the connection of lightweight information; TCP/IP has also been used to regulate information transmission and practice social calculations. In order to come up with a respectable game plan, it is first necessary to consider a model that can identify anomalies in the Internet of Things and take into account recent Internet developments.

DISADVANTAGES :

- The models are generally computationally expensive to train and deploy. This may be a limitation for organizations with limited resources.
- This models require large amounts of training data to achieve good performance. This may be a limitation for organizations that do not have access to large datasets of network traffic data.

III. PROPOSED SYSTEM

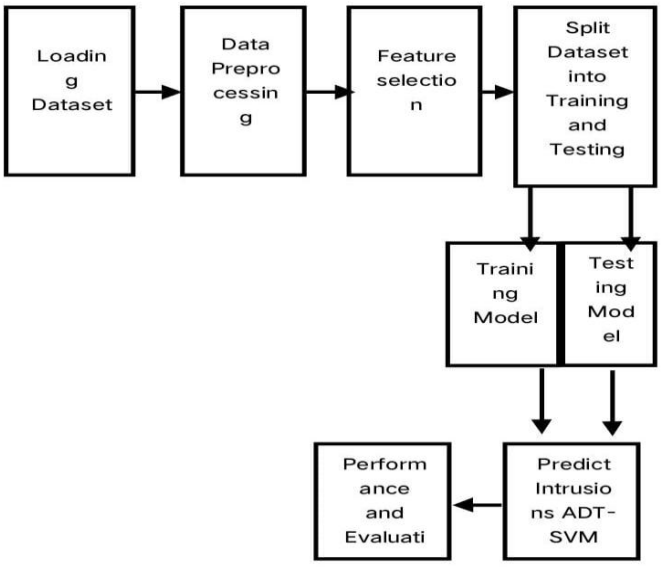
The suggested system incorporates sophisticated intrusion detection algorithms into the changing cybersecurity landscape. The system provides a

comprehensive approach to detecting potential security threats by combining a Probability Model for baseline behavior analysis, a Link-Anomaly Score computation for identifying suspicious network connections, Change Point Analysis and Dynamic Time Warping for detecting shifts in statistical properties and temporal patterns, and the Adaptive Decision Tree-Support Vector Machine (ADT-SVM) algorithm for accurate classification. Using these components, the proposed system intends to improve the adaptability and efficacy of intrusion detection, hence offering a strong defensive mechanism against emerging cyber threats. The ADT-SVM technique has the capacity to learn and categorize a variety of data properties, and the implementation process involves using the KDD dataset as a benchmark to assess the system's performance. plays a critical part in the proposed system, resulting in a more robust and responsive cybersecurity architecture.

ADVANTAGES :

- Detect network intrusion with high accuracy.
- Focus on the social aspect of the posts. This allows you to detect emerging topics even if they are not well-defined or if they are not using specific keywords. For example, if a new topic is emerging about a natural disaster, you may be able to detect it by looking at the sudden increase in mentions of relevant terms such as "earthquake" and "hurricane".
- Use of a probability model. This allows you to identify posts that are anomalous, i.e., that are more likely to be about an emerging topic. This is more effective than simply counting the number of mentions of a term, as it takes into account the context of the mentions.
- Use of change point analysis. This allows you to identify the time points at which new topics are emerging. This is important because it allows you to track the evolution of topics over time and to identify new trends early on.

ARCHITECTURE DIAGRAM



4.1 Probability Model

This lesson focuses on the creation and use of a probability model for interpreting network data. The probability model most likely evaluates the possibility of specific events or patterns in the data, offering a basic knowledge of the baseline behaviour. By creating a probability distribution, anomalies may be found by deviating from predicted patterns, allowing the system to detect possibly malicious activity.

4.2 Calculating the Link-Anomaly Score:

In this module, the system computes link anomaly scores to measure the irregularity of network links or connections. The calculation entails examining several properties related with network connections, such as traffic patterns, communication frequencies, and data transfer volumes. A higher link-anomaly score may suggest suspicious or anomalous behaviour, alerting the intrusion detection system to possible security concerns within the network.

4.3 Change Point Analysis and DTO

These subject covers change point analysis and Dynamic Time Warping (DTO) methodologies. Change point analysis seeks to uncover changes or variations in the statistical features of data, which may indicate possible security events. DTO, on the other hand, includes assessing sequence similarity across time to help in the discovery of temporal patterns. Integrating these strategies improves the

system's capacity to adapt to changing cyber threats and detect abnormal activity.

4.4 ADT-SVM Detection Method

The ADT-SVM Detection Method module applies the Adaptive Decision Tree-Support Vector Machine (ADT-SVM) algorithm to intrusion detection. This technique combines the flexibility of decision trees with the classification capability of support vector machines. The ADT-SVM model is trained using labelled data to discriminate between normal and abnormal network behavior. Once trained, it is used to classify incoming data properties into preset categories such as Basic, Content, Traffic, and Host, making it easier to identify possible security concerns on the network. The module will most likely include fine-tuning and improving the ADT-SVM settings to achieve optimal detection performance.

V.CONCLUSION

To summarize, the presented cybersecurity framework, which includes modules such as the Probability Model, Link-Anomaly Score computation, Change Point Analysis with Dynamic Time Warping, and the Adaptive Decision Tree-Support Vector Machine (ADT-SVM) algorithm, forms a comprehensive and adaptive intrusion detection system. This system successfully identifies possible security vulnerabilities by tackling the changing difficulties of the cyber threat landscape using probabilistic analysis, anomaly scoring, and machine learning. The incorporation of modern methodologies, as well as the use of the ADT-SVM algorithm, let the system adapt to and learn from developing cyber threats. The suggested paradigm not only provides a multifaceted approach to intrusion detection, but it also underlines the significance of continuous adaptation in response to growing cybersecurity threats.

VI. FUTURE WORK :

Future work in this sector might concentrate on improving and expanding the suggested cybersecurity framework to handle new concerns. Further investigation of sophisticated machine learning models other than ADT-SVM might improve the system's detecting capabilities. Investigating the convergence of threat information feeds with real-time network monitoring technologies might help create a more proactive protection mechanism. Furthermore, including methods for self-learning and adaptability to new attack vectors would be critical for staying ahead of changing threats.

REFERENCES

[1] Intrusion detection systems in the Internet of Things: A thorough research, S. Hajiheidari, K. Wakil, M. Badri, and N. J. Navimipour, Comput. Netw., vol. 160, pp. 165–191, Sep. 2021.

[2] "Ransomware detection and mitigation using software-defined networking: The case of WannaCry," by M. Akbanov, V. G. Vassilakis, and M. D. Logothetis, published in Computer Science and Electrical Engineering, vol. 76, pp. 111–121, June 2020.

[3] "A semi-boosted nested model with sensitivity-based weighted binarization for multi-domain network intrusion detection," by J. W. Mikhail, J. M. Fossaceca, and R. Iammartino Published May 27, 2019, in ACM Trans. Intell. Syst. Technol., vol. 10, no. 3.

[4] "Improved PSO AdaBoost ensemble algorithm for imbalanced data," by K. Li, G. Zhou, J. Zhai, F. Li, and M. Shao Sensors, March 2022, vol. 19, no. 6, p. 1476 In Proc. IEEE

[5] SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov.(SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Aug. 2021, pp. 1400–1405, J. Liu and S. S. Chung, "Automatic feature extraction and selection for machine learning based intrusion detection,"

[6] Anomaly network-based intrusion detection system employing a dependable hybrid artificial bee colony and AdaBoost algorithms, by M. Mazini, B. Shirazi, and I. Mahdavi, published in J. King Saud Univ.-Comput. Inf. Sci., vol. 31, no. 4, pp. 541–553, Oct. 2020.

[7] The evolution of Ethernet Passive Optical Network (EPON) and future trends, Obite, E. T. Jaja, G. Ijeomah, and K. I. Jahun, Optik, vol. 167, pp. 103–120, Aug. 2021.

[8] D. Yoo, P. D., Kim, K., H. C. Tanuwidjaja, E. Aminanto, R. Choi, and Weighted feature selection for Wi-Fi impersonator detection, deep abstraction, March 2021, 621–636 in IEEE Trans. Inf. Forensics Security, vol. 13, no. 3.

[9] "Classification by pairwise coupling of imprecise probabilities," B. Quost and S. Destercke, Pattern Recognition, vol. 77, pp. 412–425, May 2020.

[10] I-SIamids: an enhanced siam-IDS for managing class imbalance in network-based intrusion detection systems, P. Bedi, N. Gupta, and V. Jindal, Appl. Intell. (2021