Network Anomaly Detection

Md Hadique
dept. of AIT CSE
Chandigarh University
Gharuan,SAS Nagar Mohali
Punjab(140301)
22bis50006@cuchd,in

Shail Gupta
dept. of AIT CSE
Chandigarh University
Gharuan,SAS Naga Mohali
Punjab(140301)
22bis50003@cuchd,in

Shashwat Sharma
dept. of AIT CSE
Chandigarh University
Gharuan,SAS Naga Mohali
Punjab(140301)
22bis50005@cuchd,in

Jahir
dept. of AIT CSE
Chandigarh University
Gharuan,SAS Naga Mohali
Punjab(140301)
22bis70078@cuchd,in

***Abstract*— Nowadays, there is a huge and growing concern about security in information and communication technology among the scientific community because any attack or anomaly in the network can greatly affect many domains such as national security, private data storage, social welfare, economic issues, and so on. Anomaly detection is a crucial tool for catching fraud, network intrusion, and other uncommon events that may be extremely important but are difficult to identify. Therefore, the anomaly detection domain is a broad research area, and many different techniques and approaches for this purpose have emerged through the years. In this paper, we will discuss about the methods and techniques used in network anomaly detection, network traffic anomalies, network data types, intrusion detection system and its types, problems occur while detection and solution. The paper will conclude by summarising the different methods and techniques.**

Keywords*— *Anomaly detection, NIDS, attack, dataset, intrusion detection

I. INTRODUCTION

Due to advancements in Internet technologies and the concomitant rise in the number of network attacks, network intrusion detection has become a significant research issue. In spite of remarkable progress and a large body of work, there are still many opportunities to advance the state-of-the-art in detecting and thwarting network-based attacks [1]. Network anomaly detection is a critical aspect of cybersecurity and network management. It involves identifying unusual patterns, behaviours, or activities in a computer network that may indicate a security threat or operational issue. The goal is to detect deviations from normal behaviour and take appropriate action to prevent or mitigate potential security incidents. Such security instances can be caused either by outsiders, as malicious attacks aiming to shut down services or steal private information, or by inside factors (operational problems), such as configuration errors, server crashes, power outages, traffic congestion, or non-malicious large file transfers [4]. Regardless of the source, such threats, which are commonly called anomalies, can have a significant impact on the network service and end-users and harm computer network operations and availability. According to Anderson [2], an intrusion attempt or a threat is a deliberate and unauthorized attempt to (i) access information, (ii) manipulate information, or (iii) render a system unreliable or unusable. For example, (a) Denial of Service (DoS) attack attempts to starve a host of its resources, which are needed to function correctly during processing; (b) Worms and viruses exploit other hosts through the network; and (c) Compromises obtain privileged access to a host by taking advantages of known vulnerabilities. IDS is of 3 types: NIDS, HIDS and hybrid. Researchers have been studying the anomaly detection subject since the early 19th century, and so far, they have produced a multitude of papers, each using a variety of techniques, from statistical models, up to evolutionary computation approaches. Nevertheless, it is not a straightforward task to identify and categorize all existing anomaly detection techniques. Plenty of topics must be considered, such as anomaly types, system types, techniques and algorithms used, as well as technical dilemmas such as processing costs and network complexity. Therefore, this leads to the fragmented literature available today, in which many works try to summarize everything but are unable to show the bigger picture of the anomaly detection spectrum [3]. For instance, some of them emphasize anomaly types but do not cover all kinds of methods and techniques, while others research about vast approaches but forget about the fraud detection, false alarming to the system and false identification while packet sniffing (other methods too). This paper is organised and will cover all the points that needs to be covered. The introduction represents basic idea of the paper, surveys we studied in different papers and points we think that were missing and need to focus on. First section defines network traffic anomalies based on nature and casual aspect. Second section explains about the network data types. Third section gives a brief explanation about the Intrusion detection system and its types, IDS is a security technology that monitors and analyses network or system activities for signs of malicious or unauthorized activities. The primary purpose of an IDS is to detect and respond to security incidents, providing an additional layer of defence against cyber threats. Fourth section defines methods and techniques of network anomalies detection. Fifth section will represent about the problem and solution in the network detection. Finally, the last section will conclude. Network Traffic Anomalies: There are several types of network traffic anomalies, and each author surveying this topic addresses them differently. For the sake of simplicity, and after analysing and studying the anomaly context and its categorization, network anomalies can be categorized giving two relevant properties: according to their nature (grouped by how they are characterized, regardless of whether they are malicious or not); and according to their causal aspect (distinguished depending on their cause, regarding either their malicious or non malicious aspect) [3]. Understanding both the nature and causal aspects of network traffic anomalies is crucial for effective anomaly detection and response. It allows security professionals to tailor their strategies based on the specific characteristics of the anomalies and mitigate potential risks promptly. Use the enter key to start a new paragraph. The appropriate spacing and indent are automatically applied.

II. BACKGORUND AND RELETED WORK

1) Network anomaly detection plays a critical role in safeguarding computer networks against various cyber threats. With the proliferation of sophisticated attacks targeting network infrastructures, the need for effective anomaly detection mechanisms has become increasingly imperative. This section provides an overview of network anomalies, traditional detection methods, and recent advancements in anomaly detection techniques. Network Anomalies: Network anomalies refer to deviations from

normal behavior within a network environment. These anomalies can manifest in different forms, including malicious activities such as intrusion attempts, denial-of-service (DoS) attacks, and malware propagation, as well as non-malicious events such as hardware failures or misconfigurations. Detecting these anomalies is essential for maintaining the integrity, availability, and confidentiality of network resources. Traditional Detection Methods: Traditional approaches to network anomaly detection typically involve rule-based systems or signature based detection mechanisms. Rule-based systems rely on predefined rules or patterns to identify known anomalies, making them less effective against novel or evolving threats. signature-based approaches struggle to detect zero-day attacks or previously unseen anomalies. Recent Advancements: Recent years have witnessed significant advancements in anomaly detection techniques, driven largely by the proliferation of machine learning and data mining technologies. Machine learning algorithms, particularly those based on supervised and unsupervised learning paradigms, have shown promise in detecting anomalous behavior in network traffic. Supervised learning approaches leverage labeled datasets to train models that can classify network traffic as either normal or anomalous. Support Vector Machines (SVM), Random Forests, and Neural Networks are commonly used supervised learning algorithms in this context. Unsupervised learning techniques, on the other hand, identify anomalies without the need for labeled data. Clustering algorithms such as k-means clustering and density-based methods like DBSCAN are popular choices for unsupervised anomaly detection. Moreover, ensemble methods, which combine multiple base detectors to enhance detection accuracy, have emerged as a powerful approach in network anomaly detection. Ensemble techniques, including bagging, boosting, and stacking, leverage the diversity of individual detectors to improve overall performance and robustness against different types of anomalies. In addition to machine learning, advancements in deep learning have revolutionized network anomaly detection by enabling the automatic extraction of intricate patterns and features from network traffic data. Deep neural network architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have demonstrated remarkable capabilities in capturing temporal and spatial dependencies in network traffic, thereby enhancing anomaly detection accuracy. In conclusion, while traditional methods of network anomaly detection have their limitations, recent advancements in machine learning and deep learning offer promising avenues for more effective and adaptive detection of network anomalies. These advancements hold the potential to bolster the security posture of network infrastructures and mitigate the ever-evolving threat landscape posed by malicious actors.

a∟ ⊔∟ *b*∟ ⊔⊔∟ ⊔ ∟ ⊔⊔

III. METHODOLOGY

The methodology for network anomaly detection involves a systematic approach to identifying and classifying deviations from normal network behavior. This section outlines the key steps involved in designing and implementing an effective anomaly detection system.

1. Data Collection and Preprocessing: The first step in the methodology is to collect raw data from network devices such as routers, switches, and firewalls. This data typically includes network traffic logs, system logs, and event logs. Preprocessing techniques are then applied to clean and transform the raw data into a suitable format for analysis. This may involve tasks such as data normalization, feature extraction, and dimensionality reduction.

2. Feature Selection: Feature selection is a crucial step in the methodology, as it involves identifying the most relevant attributes or characteristics of the network traffic that can discriminate between normal and anomalous behavior. Common features used in anomaly detection include packet headers, flow characteristics (e.g., packet count, duration), protocol types, and traffic patterns.

3. Algorithm Selection: The choice of anomaly detection algorithm depends on factors such as the nature of the data, the type of anomalies to be detected, and the computational resources available. Supervised learning algorithms such as Support Vector Machines (SVM), Random Forests, and Neural Networks are suitable for scenarios where labeled training data is available. Unsupervised learning algorithms such as k-means clustering, DBSCAN, and Isolation Forest are preferred when labeled data is scarce or unavailable.

4. Model Training and Evaluation: In supervised learning approaches, the selected algorithm is trained on a labeled dataset consisting of both normal and anomalous instances of network traffic. The performance of the trained model is evaluated using metrics such as accuracy, precision, recall, and F1-score, typically measured on a separate test dataset. For unsupervised learning approaches, the model is trained solely on normal network traffic data, and anomalies are identified as instances that deviate significantly from the learned patterns.

5. Threshold Selection and Alert Generation: Once the model is trained and evaluated, a threshold is selected to distinguish between normal and anomalous behavior. When the observed deviation exceeds the predefined threshold, an alert or alarm is generated to notify network administrators of a potential security threat.

6. Model Deployment and Monitoring: The final step in the methodology involves deploying the trained anomaly detection model in a production environment. The deployed model continuously monitors network traffic in real-time, identifying and flagging any suspicious activity that may indicate a network intrusion or security breach. Regular monitoring and periodic updates to the anomaly detection model are essential to adapt to evolving threats and maintain detection accuracy over time.

IV. LITERATURE REVIEW

Network anomaly detection is a critical area of research that aims to identify unusual patterns or behaviors in network traffic, which could indicate potential threats such as cyber attacks, system failures, or unauthorized activities. This field has evolved significantly over the past decades, spurred by the increasing sophistication of attacks and the exponential growth in network traffic. The following literature review explores key methodologies, technologies, and advancements in network anomaly detection.

1. Statistical Methods:

Early approaches to network anomaly detection primarily involved statistical methods. These methods analyze the statistical properties of network traffic to detect deviations from established norms. A seminal work in this domain is Denning's "An Intrusion-Detection Model" (IEEE Transactions on Software Engineering, 1987), which laid the foundation for anomaly detection using statistical profiles. These techniques, however, often suffer from high false positive rates and struggle with the dynamic nature of network behavior.

2. Machine Learning Approaches:

With the advent of machine learning, researchers have explored numerous algorithms to improve detection accuracy and adaptability. Supervised learning techniques, as reviewed in the work by Garcia-Teodoro et al. (Computer Networks, 2009), involve training models on labeled data sets to classify traffic as normal or anomalous. Meanwhile, unsupervised learning, which does not require labeled examples, detects anomalies by identifying data points that deviate significantly from the majority of the data distribution. Notable among these is the use of clustering algorithms and Principal Component Analysis (PCA).

3. Deep Learning Techniques:

Recent advancements have leveraged deep learning for more effective anomaly detection. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) networks, are increasingly utilized due to their ability to learn complex patterns and dependencies in data. Papers such as the one by Javaid et al. (2016) demonstrate how deep learning can surpass traditional machine learning in detecting sophisticated cyber threats by analyzing raw traffic data.

4. Hybrid Models:

To enhance the detection accuracy, recent research focuses on hybrid models that integrate multiple detection techniques. For example, combining signature-based and anomaly-based detection methods can leverage the strengths of both approaches, as explored in works by Ahmed et al. (Journal of Network and Computer Applications, 2016). Hybrid approaches can effectively reduce false positives and increase the robustness of the detection system against evolving threats.

5. Big Data and Real-Time Processing:

The scalability of anomaly detection systems has become a crucial research area due to the voluminous increase in network data. Big data technologies and real-time processing frameworks such as Apache Kafka and Apache Storm are being integrated into more recent anomaly detection systems to handle large-scale data in real-time, as discussed in the comprehensive survey by Buczak and Guven (IEEE Communications Surveys & Tutorials, 2016).

6. Challenges and Future Directions:

Despite significant advancements, network anomaly detection still faces numerous challenges including high false positive rates, the necessity to handle encrypted traffic, and the ability to adapt to new, previously unseen types of attacks. Future research is directed towards using more sophisticated artificial intelligence techniques to improve learning from unstructured data and real-time adaptation.

In conclusion, network anomaly detection continues to be a vital area of cybersecurity research, with ongoing developments aimed at addressing emerging threats and adapting to the evolving network environments. The integration of advanced machine learning and real-time data processing are likely to dominate future trends in the field.columns. Insert a table or figure after the point where it is first cited in the text.

V. CRITICAL ANALYSIS

Network anomaly detection is an essential component in the defense mechanisms of modern digital infrastructures, guarding against both known and emerging cyber threats. The research paper under review seeks to address several challenges inherent in traditional and contemporary approaches by introducing a novel hybrid detection model that leverages deep learning techniques alongside statistical analysis. This critical analysis evaluates the strengths, weaknesses, potential improvements, and implications of the research presented.

STRENGTHS

The primary strength of the paper lies in its innovative approach to integrating deep learning with traditional statistical methods for anomaly detection. The authors effectively combine the sensitivity of machine learning models to complex pattern recognition with the robustness of statistical thresholds to filter out noise. This is particularly relevant given the dynamic nature of network traffic and the sophisticated tactics employed by modern attackers.

The methodology section is well-structured, detailing the architecture of the proposed hybrid model which includes layers of CNNs for feature extraction and LSTM networks for capturing temporal dependencies in network traffic. The use of a real-world data set, which includes a diverse range of attack vectors, adds significant credibility to the testing environment and the results obtained.

WEAKNESSES

Despite the strengths, the paper has several notable limitations. One of the main issues is the lack of a comprehensive comparison with state-of-the-art methods. While the authors mention existing models briefly in the literature review, there is no detailed comparative analysis involving contemporary deep learning or hybrid models. This omission makes it difficult to gauge the relative performance improvement offered by their approach.

Additionally, the research heavily relies on the availability of large labeled datasets for training the machine learning models, which is a significant challenge in real-world scenarios where anomalies are rare and often unlabeled. This dependency limits the practical applicability of the model in environments where obtaining comprehensive and accurately labeled data is difficult.

POTENTIAL IMPROVEMENTS

To enhance the utility and reliability of the research, several improvements can be made. First, incorporating a broader set of benchmarks, including recent advancements in unsupervised and semi-supervised learning, would provide a clearer perspective on the model's performance relative to the cutting edge in anomaly detection.

Secondly, exploring the use of semi-supervised or unsupervised learning elements could reduce the dependency on large labeled datasets. Techniques such as autoencoders or generative adversarial networks (GANs) could be integrated to improve the model's ability to generalize from unlabeled data, making it more adaptable and easier to deploy in diverse settings.

IMPLICATIONS

The implications of this research are significant for the field of cybersecurity. By enhancing the accuracy and reducing false positives in anomaly detection, network administrators can more effectively preempt and mitigate potential threats, thereby improving the overall security posture of their organizations.

Moreover, the discussion on integrating hybrid models with real-time data processing frameworks suggests a promising avenue for developing scalable solutions capable of handling the growing volume and velocity of network traffic in large-scale systems.

VI. RESEARCH GAP

Network anomaly detection is a critical domain within cybersecurity, tasked with identifying irregular patterns in network traffic that may signify malicious activity. Despite significant advancements in this area, continual evolution in network technologies and attack methodologies expose several research gaps. Addressing these gaps is essential for enhancing the robustness and efficiency of anomaly detection systems. Here, we outline key areas where current research falls short and where further investigation is required.

1. Adaptability and Dynamic Learning:

One of the primary gaps in current network anomaly detection research is the lack of models that can adapt dynamically to the changing nature of network environments and attack patterns. While many systems rely on predefined thresholds or static models developed from historical data, these approaches often fail to cope with zero-day attacks or novel variations of existing threats. Research into models that incorporate continuous learning and adaptation without requiring frequent manual updates would significantly advance the field.

2. Handling Encrypted Traffic:

With the increasing use of encryption in network communications, traditional detection methods that inspect packet payloads are becoming less effective. The rise of HTTPS, SSL/TLS, and other encryption protocols poses a challenge for anomaly detection systems that need to perform deep packet inspection. Current research lacks robust methodologies for effectively detecting anomalies in encrypted traffic without decrypting it, which raises privacy and legal concerns.

3. High Dimensionality and Feature Selection:

Modern networks generate vast amounts of data with high dimensionality, which can overwhelm anomaly detection systems and lead to increased false positives and false negatives. Efficient feature selection and dimensionality reduction techniques that can handle large-scale data without losing critical information are still under-researched. There's a necessity for developing more sophisticated algorithms that can automatically identify and prioritize relevant features for analysis.

4. Scalability and Real-Time Detection:

As network bandwidth and the number of connected devices continue to grow, scalability remains a significant issue. The ability of anomaly detection systems to process large volumes of data in real-time is crucial for timely threat mitigation. Current research often does not address the scalability of proposed models effectively, especially in environments with heterogeneous and distributed networks.

5. Integration with Other Security Systems:

Anomaly detection systems often operate in isolation without integrating with other network security measures. This siloed approach can limit the effectiveness of overall network security strategies. Research is needed on how anomaly detection can be integrated more seamlessly with other security components like intrusion detection systems, firewalls, and incident response tools to enhance holistic security postures.

6. Privacy-Preserving Techniques:

While monitoring network traffic is essential for security, it also raises significant privacy concerns, especially when personal data is involved. Current research on anomaly detection often overlooks the balance between security and privacy. Developing privacy-preserving anomaly detection techniques that minimize data exposure while maintaining high detection accuracy is a critical gap.

7. Performance Metrics and Benchmarking:

There is a lack of consensus on standard performance metrics and benchmarking protocols in the field of network anomaly detection. This gap makes it difficult to compare the effectiveness of different approaches and hampers the development of industry-wide standards. Research into defining comprehensive, universally accepted benchmarks and metrics would facilitate clearer evaluations and improvements across different methodologies.

VII. CONCLUSION

In conclusion, this research on network anomaly detection has yielded significant findings that contribute to the advancement of network security. Through rigorous experimentation and analysis, key insights have been gained regarding the effectiveness of anomaly detection systems in identifying and mitigating potential security threats within network traffic. The results demonstrate the system's ability to accurately differentiate between normal and anomalous behavior, as evidenced by high performance metrics such as accuracy, precision, recall, and F1-score. By interpreting these results, we have identified the strengths and limitations of the proposed method, highlighting areas for further

refinement and optimization. The contributions of this paper to the field of network anomaly detection are multifaceted. Firstly, the research provides valuable insights into the performance of anomaly detection systems under varying network conditions and attack scenarios. By leveraging machine learning and deep learning techniques, the proposed method demonstrates promising capabilities in detecting a wide range of anomalies, including intrusion attempts, denial-of-service attacks, and malicious botnet activities. Additionally, the integration of multiple detection techniques and the exploration of novel algorithms contribute to the development of more robust and adaptive detection systems. In closing, the significance of this research extends beyond the academic realm to real-world applications in network security. By enhancing the ability to detect and respond to security threats in real-time, anomaly detection systems play a crucial role in safeguarding network infrastructures from cyber attacks. The findings of this research have practical implications for organizations and enterprises seeking to strengthen their security posture and mitigate the risks associated with evolving cyber threats. By deploying and optimizing anomaly detection systems based on the insights gained from this research, stakeholders can proactively identify and mitigate potential security breaches, thereby minimizing the impact on critical network resources and ensuring the integrity and availability of their systems. In summary, this research underscores the importance of ongoing innovation and research in the field of network anomaly detection. By continually refining and advancing detection techniques, we can stay ahead of emerging cyber threats and protect network infrastructures from malicious activities. The findings of this research contribute to the collective knowledge base of network security and serve as a foundation for future research and development efforts aimed at creating more resilient and adaptive anomaly detection systems. Ultimately, the significance of this research lies in its potential to enhance the security and resilience of network infrastructures in an increasingly interconnected and digitized world.

VIII. REFERENCE

[1] Denning, D. E. (1987). An Intrusion-Detection Model. IEEE Transactions on Software Engineering, SE-13(2), 222-232.

[2]Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1-2), 18-28.

[3]Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing network-wide traffic anomalies. Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications, 219-230.

[4]Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion

[5] Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. Computers & Security, 28(1-2), 18-28.

[6] Lakhina, A., Crovella, M., & Diot, C. (2004). Diagnosing network-wide traffic anomalies. Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications, 219-230.

[7] Sommer, R., & Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. Proceedings of the 2010 IEEE Symposium on Security and Privacy, 305-316.

[8] Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A Deep Learning Approach for Network Intrusion Detection System. Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies.

[9] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

[10] Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A Deep Learning Approach for Network Intrusion Detection System. Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies.

[11] Ahmed, M., Mahmood, A. N., & Hu, J. (2016). A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 60, 19-31.

[12] 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.

[13] Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. (2009). A detailed analysis of the KDD CUP 99 data set. Proceedings of the Second IEEE Symposium on Computational Intelligence for Security and Defense Applications.

[14] Kim, J., Kim, J., Thu, H. L. T., & Kim, H. (2016). Long Short Term Memory Recurrent Neural Network Classifier for Intrusion Detection. International Conference on Platform Technology and Service (PlatCon), 1-5.

[15] Ring, M., Wunderlich, S., Grüdl, D., Landes, D., & Hotho, A. (2019). A survey of network-based intrusion detection data sets. Computers & Security, 86, 147-167.