Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 10, No.1, October : 2024

AN INTELLIGENT AI FRAMEWORK FOR DETECTING ANOMALIES IN NETWORK TRAFFIC USING SWARM INTELLIGENCE AND MACHINE LEARNING

Mr Panjatcharam V G Assistant Professor, Department of Computer Science, VET Institute of Arts and Science (Co-ed) College, Erode
Dr. Selvanayaki K Assistant Professor, Department of Computer Science and Applications, VET Institute of Arts and Science (Co-ed) College, Erode

Abstract
This paper offers a comprehensive review of the current research landscape and recent advancements in network anomaly detection, focusing on the integration of swarm intelligence and machine learning methods. We begin by outlining the fundamental concepts, common techniques, and challenges related to network traffic analysis through various methods, including swarm intelligence approaches. Inspired by the collective behaviour of social organisms such as ants, bees, and birds, swarm intelligence is increasingly utilized in network security to enhance threat detection, response, and overall system resilience. We provide an in-depth summary of the leading anomaly detection techniques, covering statistical methods, machine learning algorithms, deep learning approaches, and behavioral analysis. This section examines their core principles, highlights notable studies, and discusses their strengths, weaknesses, and appropriate application scenarios.Next, we explore hybrid anomaly detection methods, detailing their motivations, common strategies, and key representative works. We assert that hybrid approaches are a vital development avenue for improving anomaly detection capabilities. The paper also evaluates the practical effectiveness of various methods in real-world network security tasks, presenting a quantitative comparison in tabular format.Finally, we look ahead to future trends in network anomaly detection using swarm intelligence, emphasizing objectives such as increased intelligence, automation, federated learning, and interpretability. We address challenges in anomaly detection, including data heterogeneity, the complexity of security threats, model robustness, privacy concerns, and interpretability. We argue that overcoming these challenges will require interdisciplinary approaches, enhanced governance of security big data, and a shift from passive defense strategies to proactive immunity. As the importance of cybersecurity grows, along with the influence of disruptive technologies like big data, artificial intelligence, and blockchain, network anomaly detection is set to face significant new opportunities and challenges.

Keywords:
Pre-processing, Feature Extraction, Feature Selection, Classification, Swarm Intelligence,Particle Swarm Optimiuzation , Machine Learning Algorithms.

1 Introduction

From the early days of computing to the present, the need for automated hardware and software solutions to protect data, files, documents, and other information stored on computers has become increasingly apparent. This is especially true for shared systems and systems that can be accessed over the Internet. The collection of tools designed to protect data and thwart hackers falls under the umbrella of "computer security. The introduction of distributed systems and the use of networks and communication facilities to transmit data between users and computers, or between computers, have heightened the need for robust network security measures. Network security involves protecting data during transmission and is generally synonymous with Internet security. It encompasses measures that provide access control and protection for our data, ensuring safe and secure communication across the network.

Network security begins with authentication, typically using a username and password. It encompasses the measures and policies implemented by network administrators to prevent and monitor unauthorized access, system modifications, misuse, or denial of computer networks and their accessible resources.

Essentially, network security involves managing who can access data within a network, with control resting in the hands of network administrators. This is crucial for both individual users and organizations.

Firewalls are used to enforce access policies, determining which services network users can access. However, firewalls may not always detect potentially harmful content like computer worms or Trojans transmitted over the network. To address this, anti-virus software and intrusion detection systems (IDS) are employed to identify and manage malware. Additionally, tools such as Wireshark can monitor network traffic, log anomalies, and provide data for auditing and high-level analysis. Encryption is also used to protect communication between hosts, ensuring privacy and security.

1.1 Fundamentals of Network Security

System and network technology is crucial for a wide range of applications, with security being a fundamental concern for networks and their associated applications. Despite its importance, there is a noticeable gap between the developers of security technologies and those who design networks. Network design benefits from the well-established Open Systems Interconnection (OSI) model, which provides modularity, flexibility, ease of use, and standardized protocols. This model allows for the easy combination of different layer protocols into stacks, supports modular development, and enables updates to individual layers without affecting others.

In contrast, secure network design lacks a similarly developed methodology. Unlike network design, secure network design does not offer the same level of standardization or modularity, making it challenging to manage security complexities effectively.When addressing network security, it is crucial to ensure the security of the entire network, not just the computers at each end of the communication chain. The communication channel itself must be protected against potential attacks. A hacker could exploit vulnerabilities in the communication channel to intercept data, decrypt it, and insert false information. Therefore, securing the network infrastructure **is** as vital as safeguarding the end computers and encrypting messages.

1.2 Literature Survey

Network security has become increasingly vital for personal computer users, organizations, and the military. With the rise of the internet, security concerns have escalated, making the history of security essential for understanding the development of security technologies. Modifying the internet's architecture can help minimize potential attacks. By understanding various attack methods, appropriate security measures can be developed. Many businesses protect themselves from internet threats using firewalls and encryption, while also creating an "intranet" to stay connected to the internet while safeguarding their internal networks. The field of network security is extensive and continuously evolving. To grasp current research, it is crucial to have a background in the internet, its vulnerabilities, attack methods, and security technologies. Shi-Jinn Horng *et al.,* in [1] designed a new flow for intrusion detection system using Support Vector Machine (SVM) technique. The famous KDD Cup 1999 dataset was used to evaluate the proposed system. Compared with other intrusion detection systems that are based on the same dataset, this system exhibited better performance in the detection of DoS and Probe attacks, and the best performance in overall accuracy. Mohammad Wazid in [2] has used hybrid anomaly detection technique with the k-means clustering. WSN are simulated using Optimized Network Engineering Tool (OPNET) simulator and the resultant dataset consists of traffic data with end to end delay data which has been clustered using WEKA 3.6.

information of source and destination. Dat Tran *et al.,* [10] proposed Fuzzy Gaussian mixture modeling method for network anomaly detection. It was a mixture of Gaussian distributions used to represent the network data in multi-dimensional feature space. Using fuzzy C-means estimation, Gaussian parameters were estimated and the whole work is carried out with the KDD Cup data set. The proposed method produced here is more effective than the vector quantization method.Vahid Golmah in [11] developed a hybrid technique using C5.0 and SVM algorithm to evaluate the performance of the hybrid technique with DARPA dataset. The motivation behind this

hybrid approach was to improve the accuracy of the intrusion detection system when compared to using individual SVM and C5.0. Due to the mixture of SVM and C5.0, it took less execution time. Punam Mulak in [12] has used hybrid technique In this experiment, it has been observed that two types of anomalies namely misdirection and black hole attacks were activated in the network .Shun-Sheng Wang *et al.*, [3][4] have designed an integrated intrusion detection system using intrusion dataset from UCI repository .The dataset trained well using Back Propagation Neural Network (BPNN) and the output is used as an important parameter in Adaptive Resonance Theory (ART) model to cluster the data. Finally the outputs received from both techniques are compared and the ART model provided the best accuracy rate and overall performance.Mohit Malik *et al.*, [5] applied the rule based technique for detecting the security attack in WSN. They identified ten important security attack types developed a fuzzy rule based system for calculating the impact of security attacks on the wireless sensor network .Reda M. Elbasiony *et al.*, [6] proposed a hybrid detection framework using K-means clustering algorithm to detect novel intrusions by clustering the network connections. In this hybrid framework, the anomaly part was improved by replacing the k-means algorithm with the weighted k-means algorithm. LeventKoc *et al.*, [7] proposed a new technique Hybrid Naïve Bayes (HNB) and excelled in a superior performance in terms of accuracy, error rate and misclassification cost. In early stages the traditional Naïve Bayes model are used but the result produced by HNB is better than traditional Naïve Bayes. The results they have produced indicate that this model significantly improves the accuracy for detecting the denial-of-services (DoS) attacks. Wenying Fenga *et al.*, [8] introduced a new way of combining algorithm for the better result in detecting intrusions and classified the network activities into normal or abnormal by reducing the misclassification rate. It combined Support Vector Machine method and the Clustering based on Self-Organized Ant Colony Network to take the advantages by avoiding their weaknesses. This Experiments show that CSVAC (Combining Support Vectors with Ant Colony) outperforms better the SVM or CSOACN in terms of both classification rate and run-time efficiency.Megha Bandgar *et al.*, [9] described a novel approach using Hidden Markov Models (HMM) to detect Internet attacks and described an intrusion detection system for detecting a signature based attack. They have performed single and multiple HMM model for source separation both on IP and port by combining Boundary cutting algorithm and clustering algorithm. The motivation for using this hybrid approach is to improve the accuracy of the intrusion detection system and to provide better result than other clustering.Venkata Suneetha Takkellapati in [13] proposed a new system with Information Gain (IG) and Triangle Area based KNN algorithm is for selecting more discriminative features. Then the Greedy k-means clustering algorithm was combined with SVM classifier to detect Network attacks. This system achieved a accuracy detection rate and less error rate .All these experiments were conducted in KDD CUP 1999 training data set. Vaishali Kosamkar in [14] developed technique of combining C4.5 Decision Tree and Support Vector Machine (SVM) algorithm in order to achieve high accuracy and diminish the false alarm rate. For feature selection stage, the Correlation- Based Feature Selection (CFS) algorithm was used for better accuracy result. Harmeet Kaurl in [15] designed a model to reduce the delay in the network and to produce an end to end data in good speed. A simulated WSN using SPEED protocol was used. It was concentrating on two different performance parameters throughput and energy consumption for analysis. BCO (Bee Colony Optimization) algorithm was used to give better results with high throughput and low energy consumption.H. Oh, I. Doh and K. Chae in [16], the authors proposed a real-time intrusion detection system based on the Self- Organizing Map (SOM); an unsupervised learning technique that is appropriate for anomaly detection in wireless sensor networks. The proposed system was tested using KDD'99 Intrusion Detection Evaluation dataset. The system groups similar connections together based on correlations between features. A connection may be classified as normal or attack. Attacks are classified again based on the type of attack. It took the system 0.5 seconds to decide whether a given input represents a normal behavior or an attack. N. Ye and X. Li in [17], A data mining algorithm called Clustering and

Classification Algorithm Supervised (CCA- S) was developed for intrusion detection in computer networks. The algorithm is used to learn signature patterns of both normal behaviors and attacks. Compared to anomaly detection techniques, the signature recognition techniques always produce true alarms, but not being the capability to detect unknown attacks. The algorithm's scalability and incremental learning were improved performance the decision tree algorithms. G. Singh, F. Masseglia, C. Fiot, A. Marascu and P. Poncelet in [18], the authors addressed the main drawback of detecting intrusions by means of anomaly (outliers) detection. In their work, they added a new feature to the unknown behaviors before they are considered as attacks, and they claim that the proposed system guarantees a very low ratio of false alarms, making unsupervised clustering for intrusion detection more effective, realistic and feasible. K. Faraoun and A. Boukelif in [19], a genetic programming approach for multi-category pattern classification applied to network intrusion detection, proposed to reduce the input patterns dimension towards a better inter-classes discrimination, and achieved through non-linear transformations on the original datasets. W. Lee, S. Stolfo, P. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop and J. Zhang in [20], a real time data mining based intrusion detection like accuracy, efficiency and usability in intrusion detection in real time environments. It used the artificial anomalies, multiple model and adaptive learning algorithms to address the above issues respectively. K. Ioannis, T. Dimitriou and F. C. Freiling in [21], a light weight intrusion detection scheme was proposed to identify or detect the effect of attack in WSN by utilizing the concept of collaborative communication methodology. They also formulated the general rules for the WSN too. D. Farid, J. Darmont, N. Harbi, N. Hoa and M. Rahman in [22], the authors addressed the complexity of the intrusion detection datasets, as most of them are complex and contain large number of attributes. Some of these attributes may be redundant or do not have significant contribution for intrusion detection. The aim of this work was to specify effective attributes from the training dataset to build a classifier using data mining algorithms. Experimental results on KDD'99 intrusion detection dataset show that the proposed approach achieves high classification rates and reduces false positives in such environment with limited computational resources. J. Zhang and M. Zulkernine in [23], the authors focused on the high rate of false positive in intrusion detection associated with an intent of achieving a high rate of false positives in intrusion detection, a modified random forest algorithm was developed, and tested using WEKA tool, testing was conducted on KDD CUP 99 dataset for the above said claim.M. Tavallaee, E. Bagheri, W. Lu and A. Ghorbani in [24], to overcome the short coming of KDD CUP 99 dataset, a new dataset called NSL-KDD [24] was proposed and presented a detailed statistical analysis model to evaluate the intrusion detection systems. Campose *et al.,* [25] proposed a Database Centric Architecture for Intrusion Detection (DAID) system in Oracle 10g to address the challenges in designing and implementing data mining based intrusion detection systems. DAID offered numerous advantages in terms of scheduling capabilities, alert infrastructure, data analysis tools, security, scalability, and reliability. Prothives and S. Srinoy in [26], an intrusion detection system based on Adaptive Resonance Theory (ART) and Rough Set Theory [38] to detect the known attacks and also new unknown attacks by creating new clusters using ART and RT. Güneş Kayacık, A. Nur Zincir-Heywood and M. I. Heywood in [27], a feature relevance analysis [27] was conducted on KDD CUP 99 to enlist the effects of features in detecting the intrusion in systems. mini *et al.,* in [28] introduced an intrusion detection approach based on Adaptive Resonance Theory (ART) and Principal Component Analysis (PCA). The PCA is used for feature selection to reduce the computational complexity and training time of ART. Experimental results show that modifications proposed in this approach improved the speed and accuracy of detection . Xiao and H. Song in [29], an intrusion detection system called Unsupervised Neural Net based Intrusion Detector (UNNID) was introduced to provide the facilities for training, testing, and tuning of unsupervised Adaptive Resonance Theory (ART) with neural networks used for intrusion detection. E. Skoudis in [30], to mention a few of the attacks Smurf attacks, also known as directed broadcast attacks, and are popular form of DoS

packet floods. Smurf attacks rely on directed broadcast to create a flood of traffic for a victim. The attacker sends a ping packet to the broadcast address for some network on the Internet that will accept and respond to directed broadcast messages, known as the Smurf amplifier. The attacker uses a spoofed source address of the victim. If there are 30 hosts connected to the Smurf amplifier, the attacker can cause 30 packets to be sent to the victim by sending a single packet to the Smurf amplifier. Labib and V. Rao Vemuri in [31], Neptune attacks can make memory resources too full for a victim by sending a TCP packet requesting to initiate a TCP session. This packet is part of a three-way handshake that is needed to establish a TCP connection between two hosts. The SYN flag on this packet is set to indicate that a new connection is to be established. This packet includes a spoofed source address, such that the victim is not able to finish the handshake but had allocated an amount of system memory for this connection. After sending many of these packets, the victim eventually runs out of memory resources. IPsweep and Portsweep, as their names suggest, sweep through IP addresses and port numbers for a victim network and host respectively looking for open ports that could potentially be used later in an attack.T. Eldos, M. Khubeb Siddiqui and A. Kanan in [32], author presented a contribution to the network intrusion detection process using Adaptive Resonance Theory (ART1), a type of Artificial Neural Networks (ANN) with binary input unsupervised training. they presented the feature selection using data mining techniques, towards two dimensional dataset reduction that is efficient for the initial and on-going training, and reduce the dataset both vertically and horizontally, numbers of vectors and number of features. The following Tabl1 shows the Existing representations and Search algorithmsin network security.

Table1 : Algorithms and Sample Methods in Network Security

Researchers	***Approach***	***Model representation***	***Search algorithm***
Fan et al 2000	Misuse Detection	Ordered Associative rule	FAST rule induction
Neri 2000a	Misuse Detecion	Associative rule	Genetic Algorithm
Lane 2000	Anomaly detection	Hidden Markov Model(HMM)	Baum-Welch algorithm
Dasguta et al 2001	Misuse detection	Classifier on statistics of attributes	Genetic Algorithm
Portnoy et al 2001	Anomaly detection	Outliers from clusters	Fixed width clustering
Bloedorn et al 2001	Anomaly detection	Outliers from clusters	K-mean clustering
Eskin et al 2002	Anomaly detection	Outliers from clusters	K-nearest neighbor
Staniford et al 2002	Misuse Detection	Bayes network	Simulated annealing
R. Sekar et al 2002	Specification based detection	State Machine	Specification Language
Mukkamala et al 2003	Misuse detection	Support Vector Machine (SVM)	SVM algorithm
Me-Ling Shyu et al 2003	Anomaly detection	Principle Component Ananlysis(PCA) feature reduction	Principal Component Algorithme
Gopi K. Kuchimanchi et al 2004	Misuse detection	PCA,Neural Network Feature reduction	Decision tree

Table 2 : Review on Best methods and systems in Network Security .

Sno	Author	Method	Remark
1	Shi-Jinn Horng	Intrusion detection system - Support Vector Machine (SVM) technique	The detection of DoS and Probe attacks, and the best performance in overall accuracy 96.6%
2	Mohammad Wazid	hybrid anomaly detection - k-means clustering.	The resultant dataset consists of traffic data with end to end delay data
3	Dat Tran	Fuzzy Gaussian mixture Model - fuzzy C-means estimation	This Fuzzy method effective than the vector quantization method. Produce 88.93%
4	Vahid Golmah	Hybrid technique – SVM Technique	To improve the accuracy of the intrusion detection system when compared to using individual SVM and C5.0.
5	Punam Mulak	Misdirection and black hole attacks	To improve Qualities and Produce overall Accuracy.
6	Wang	Back Propagation Neural Network (BPNN) and Adaptive Resonance Theory (ART)	The outputs are compared and the ART model provided the best accuracy rate and overall performance.
7	Mohit Malik	The rule based technique for detecting the security attack in WSN with Fuzzy.	A fuzzy rule based system for calculating the impact of security attacks on the wireless sensor network. Produce 90% accuracy.
8	Reda M. Elbasiony	Hybrid detection framework using K-means clustering algorithm	The anomaly part was improved by replacing the k-means algorithm with the weighted k-means algorithm.The average rate is more than 80%
9	LeventKoc	Hybrid Naïve Bayes (HNB)	To produce 92.3% accuracy excelled in a superior performance in terms of accuracy, error rate and misclassification cost.
10	Wenying Fenga	Support Vector Machine method and the Clustering based on Self-Organized Ant Colony Network	Experiments show that CSVAC (Combining Support Vectors with Ant Colony) outperforms better the SVM .the System Produced 99.78%

The above Table2 explains various network traffic methods and its analysis.The literature review (Table 1 & Table 2) indicates that Support Vector Machines (SVMs) combined with various methods often achieve high accuracy and overall performance. Specifically, it highlights that integrating SVMs with swarm intelligence techniques yields exceptional results in terms of accuracy and detection performance compared to earlier methods. Based on these findings, we have chosen to focus our work on combining Support Vector Machines with Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO).

1.3 Overview of System

This paper reviews the latest research progress in traffic analysis and anomaly detection in the field of network security, focusing on introducing anomaly detection methods based on statistics, machine learning, deep learning, and behaviour analysis, discussing the advantages, disadvantages, and applicable scenarios of different methods, and providing an outlook on future development trends and challenges. By sorting out and analyzing the research status of network security traffic analysis and anomaly detection techniques, this paper can provide beneficial references and insights for researchers and practitioners in related fields, promoting technological innovation and application development in this field. At the same time, this paper also contributes useful ideas and methods for solving

increasingly severe network security problem. Network traffic refers to the data flow transmitted through the network within a certain period of time, usually measured in units of data packets or bytes [3].

1.3.1 Network traffic has the following main characteristics:

(1) Large data volume: With the continuous enrichment of network applications, network traffic has shown explosive growth. According to Cisco's forecast, global IP traffic will reach 396 extra bytes per month by 2025.

(2) Diverse types: Network traffic contains various application protocols and data formats, such as HTTP, FTP, DNS, etc. Different types of traffic have different characteristics and behavior patterns.

(3) Dynamic changes: Network traffic will dynamically change with time, location, and user behavior, exhibiting complex non-stationary characteristics.

(4) Heterogeneous distribution: Network traffic is unevenly distributed across different network nodes and links, with significant heterogeneity and locality characteristics.

Basic Methods of Network Traffic Analysis mainly includes the following three basic methods:

Packet analysis: By capturing and parsing network data.

This research paper examines the latest advancements in network traffic monitoring and anomaly detection within the field of network security. It focuses on statistical, machine learning, deep learning, and behavior analysis-based methods for detecting anomalies, discussing their respective strengths, weaknesses, and suitable application scenarios. The paper also provides an outlook on future development trends and challenges. By reviewing and analyzing the current state of network traffic analysis and anomaly detection techniques, it offers valuable insights and references for researchers and practitioners. This work aims to foster technological innovation and application development in network security, while also proposing useful strategies and methods to address the growing network security challenges.

Network traffic refers to the flow of data transmitted across a network over a specific period, typically measured in data packets or bytes. It exhibits several key characteristics:

1. **Large Volume**: Network traffic has surged significantly due to the growth of network applications. Cisco forecasts that global IP traffic will reach 396 exabytes per month by 2025.
2. **Diverse Types**: Network traffic includes various application protocols and data formats, such as HTTP, FTP, and DNS, each with distinct characteristics and behavior patterns.
3. **Dynamic Nature**: Network traffic is subject to continuous changes influenced by time, location, and user behavior, displaying complex, non-stationary characteristics.
4. **Heterogeneous Distribution**: Traffic is unevenly distributed across network nodes and links, showing significant variability and locality features.

Basic Methods of Network Traffic Analysis

Network traffic analysis primarily involves the following methods:

1. **Packet Analysis**: This involves capturing and parsing network data packets to extract key information (such as source/destination IP addresses, port numbers, and protocol types) for detailed analysis. Tools commonly used for packet analysis include Wireshark and Tcpdump.
2. **Flow Statistics Analysis**: This method measures and analyzes statistical characteristics of network traffic, such as flow rate, connection count, and packet size distribution, to identify overall patterns and trends. Common statistical indicators include mean, variance, and probability distribution.
3. **Flow Behavior Analysis**: This approach models and examines the behavioral characteristics of network traffic, such as communication frequency, duration, and interaction patterns, to uncover behavioral patterns and detect anomalous events. Methods used in behavior analysis include association rule mining and sequence pattern mining.

Statistical anomaly detection in networking involves identifying unusual patterns or deviations from the expected statistical norms within network traffic. This method leverages statistical techniques to analyze traffic data and flag anomalies that may indicate potential security threats or operational issues. Here's an overview of how statistical anomaly detection works in the context of networking:

1.4 Concept and Approach

1. **Data Collection**: Gather network traffic data, which includes various metrics such as packet sizes, flow rates, connection counts, and other relevant parameters.
2. **Feature Extraction**: Extract key features from the collected data that are crucial for analysis. Common features include average packet size, flow duration, connection frequency, and data transfer rates.
3. **Modeling Normal Behavior**: Develop a statistical model that represents normal network behavior. This typically involves calculating statistical properties such as mean, variance, and distribution patterns for the features of interest. Historical data is used to establish baseline metrics.
4. **Anomaly Detection**: Continuously monitor incoming network traffic and compare it against the established statistical model. Anomalies are detected when there are significant deviations from the baseline metrics. Various statistical methods can be applied, such as:

- **Threshold-Based Methods**: Define thresholds for statistical metrics (e.g., mean + 3 standard deviations). If the observed metrics exceed these thresholds, it is flagged as an anomaly.
- **Statistical Distribution Models**: Use probability distributions (e.g., Gaussian distribution) to model normal behavior. Deviations from these distributions can indicate anomalies.
- **Time-Series Analysis**: Analyze traffic patterns over time to detect deviations from expected trends. Methods such as autoregressive integrated moving average (ARIMA) models can be used to forecast normal behavior and identify anomalies based on deviations from forecasts.

The basic principle of statistical anomaly detection is to assume that normal data follows a certain probability distribu-tion model, and then use statistical inference methods such as hypothesis testing or likelihood estimation to determine whether unknown data conforms to the known model [15].Common probability distribution models include Gaussian distribution, exponential distribution, Poisson distribution, etc., and distribution parameters can be learned from historical data using methods such as maximum likelihood estimation. Let When $p(x)$ is less than a set threshold, it can be determined that x is an anomaly point.

2. Proposed System

The proposed system contains five important stages like Preprocessing ,Enhancement ,Feature Extraction & Selection finally Classifications. Preprocessing in network security involves preparing raw network data for further analysis and model building. Effective preprocessing ensures that data is clean, consistent, and formatted in a way that enhances the performance of subsequent security analyses, such as threat detection, anomaly detection, and classification. Here are common preprocessing methods used in network security. Enhancement methods in network security are techniques and practices designed to improve the effectiveness, efficiency, and resilience of security measures. These methods can help in strengthening defenses, optimizing performance, and adapting to evolving threats.Feature extraction in network security involves deriving meaningful attributes from raw network data that can help in identifying threats, detecting anomalies, and classifying network traffic. In feature selection for network security, filter methods are used to evaluate the relevance of features independently of any learning algorithm. Classification methods in network security are used to categorize network traffic, events, or activities into predefined classes or labels. This helps in identifying, managing, and responding to different types of network threats or anomalies.The following tools and methods are going to proposed in our network security system.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 10, No.1, October : 2024

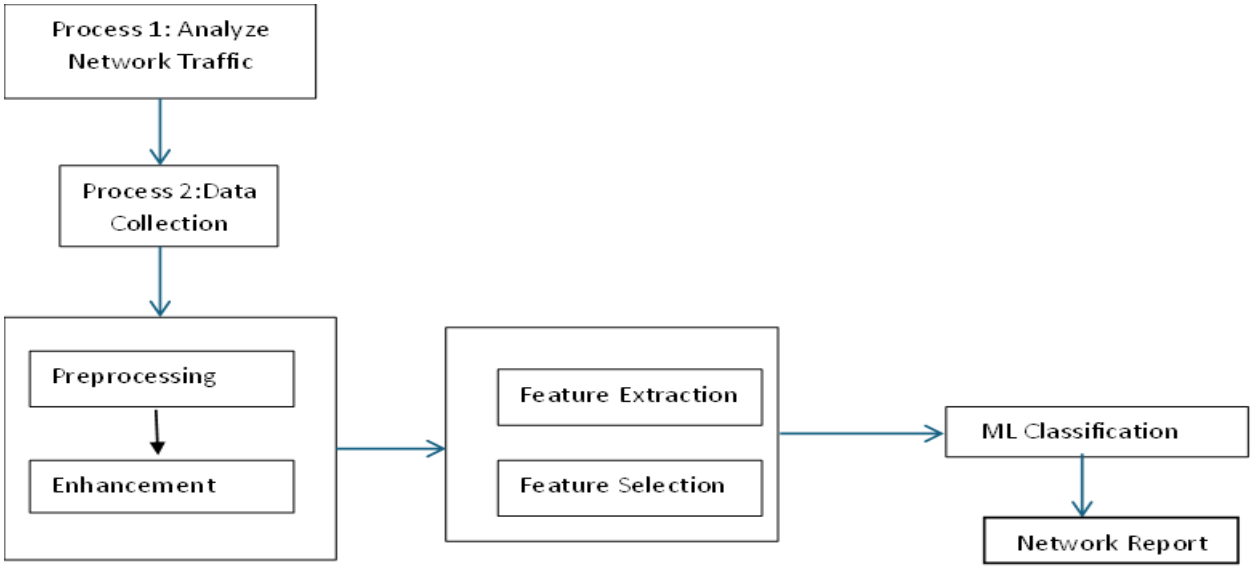

Figure 1 : Phases of Proposed System

This paper offers a comprehensive review of the current research on traffic analysis and anomaly detection within the network security domain. As network security threats become increasingly severe, there is an urgent need for more intelligent and adaptive anomaly detection techniques. Future systems should be capable of continuous learning, proactive defense, multi-domain collaboration, transparency, and control, evolving from passive defense mechanisms to active immunity.On one hand, enhancing interdisciplinary integration is crucial.

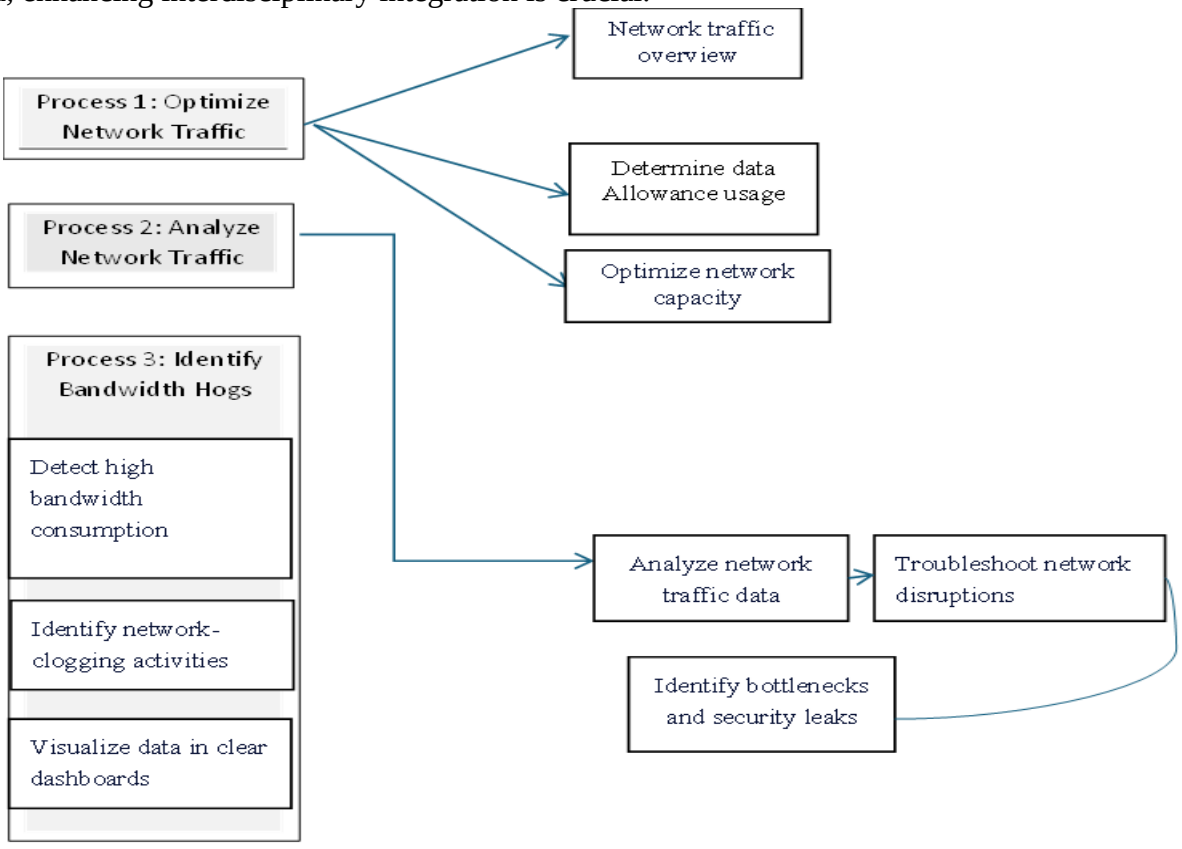

Figure 2 : Basic Process and Sub Process of Network Traffic Analysis

This includes incorporating cutting-edge theoretical advancements from fields such as artificial intelligence, game theory, and causal inference to challenge and move beyond traditional anomaly detection paradigms. On the other hand, it is essential to focus on the governance of security big data. Improving data quality, enriching data semantics, and establishing a solid foundation for intelligent

anomaly detection are key to addressing these challenges effectively. The following Table 3 shows the sample structure of proposed system for Network Traffic analysis and Table 4 describes technical SVM Algorithm.

Table 3 : Proposed Methods and Tools

Sno	Stages in Proposed Method	Methods and Techniques	Process
1	Preprocessing	Min-Max Scaling or Standardization & Normalization	Convert the network data to a common scale or range, such as transforming edge weights to a [0,1] range.
2	Enhancement	Graph-based smoothing methods and Outlier Detection	Identify and handle outliers that could skew the results.
3	Feature Extraction	Heuristic-Based Detection	To identify suspicious behavior based on known patterns or characteristics.
4	Feature Selection	Filter Methods, Correlation-Based Feature Selection,	The process of choosing the most relevant features from the extracted set to improve model performance and reduce complexity.
5	Classification	Machine learning Methods : Support Vector Machine, k-nearest neighbour (kNN), Particle Swarm Optimization(PSO), Convolutional Neural Networks (CNNs). ROC and AUC	Receiver Operating Characteristic (Roc) curve and Area Under the Curve(AUC) measure the performance of a classification model across different thresholds.

Table 4 : Technical Support Vector Machine Algorithm

The following table shows the sample steps in Support Vector Machine algorithm (SVM)

Step1 : Select some samples s randomly from the dataset D in a stratified sampling manner

Step2: Select y samples r of the same type D_a nearest to the sample s

Step 3: Select y samples t from different types of D_b

Step 4: Calculate the distance between sample s and sample r as D_{sr} and the distance between sample s and sample t as D_{st}

Conclusion

In this paper, we have reviewed various advanced techniques and methods developed by researchers for network security across WSN, IoT, Cloud Computing, WBAN, and Big Data. The article proposes future research directions focused on security threats within these domains, particularly considering the context of smart homes, where AI and multi-modal sensor technologies will be increasingly prevalent. This paper illustrates the research directions in network security. This review paper, provides a comprehensive review of network security techniques aimed at enhancing network protection. We also explored previous and innovative approaches in network security, emphasizing its crucial role in information security by safeguarding all data transmitted across networked systems. Network security encompasses the implementation of cryptographic algorithms in network protocols and applications. Ensuring the security of data has become increasingly critical.

References

[1] Shi-Jinn Horng, Ming-Yang Su, Yuan-Hsin Chen, Tzong-Wann Kao, Rong-Jian Chen, Jui- Lin Lai, Citra Dwi Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines", Elsevier Computer Network, pp.306–313, 2010.
[2] Mohammad Wazid, "Hybrid Anomaly Detection using K-Means Clustering in Wireless Sensor Networks", Center for Security, Theory and Algorithmic Research, pp. 1-17, 2014.
[3] Y.-J. Shen and M.-S. Wang, "Broadcast scheduling in wireless sensor networks using fuzzy hopfield neural network," Expert Systems with Applications, Vol. 34, No. 2, pp. 900-907, 2008.
[4] Y. Wang, M. Martonosi, and L.-S. Peh, "Predicting link quality using supervised learning in wireless sensor networks," ACM SIGMOBILE Mobile Computing and Communications Review, Vol. 11, No. 3, pp.71–83, 2007
[5] Mohit Malik, Namarta kapoor, Esh naryan, Aman Preet Singh, "Rule Based Technique detecting Security attack for Wireless Sensor network using fuzzy logic", International Journal of Advanced Research in Computer Engineering & Technology, Vol. 1, No. 4,, ISSN: 2278–1323, June 2012.
[6] Reda M. Elbasiony, Elsayed A. Sallam, Tarek E. Eltobely,Mahmoud M. Fahmy, "A hybrid network intrusion detection framework based on random forests and weighted k-means" Ain Shams Engineering Journal, vol 4, pp.753–762,2013.
[7] Levent Koc, Thomas A. Mazzuchi, Shahram Sarkani, "A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier", Elsevier, pp.13492–13500, 2012.
[8] Wenying Fenga, Qinglei Zhangc, Gongzhu Hud, Jimmy Xiangji Huange, "Mining network data for intrusion detection through combining SVMs with ant colony networks", Elsevier, pp. 127-140,2013.
[9] Megha Bandgar, Komal dhurve, Sneha Jadhav,Vicky Kayastha,Prof. T.J Parvat, "Intrusion Detection System using Hidden Markov Model (HMM)", IOSR Journal of Computer Engineering (IOSRJCE) e- ISSN: 2278-0661, p- ISSN: 2278- 8727Vol. 10, No. 3, pp. 66-70, Mar. - Apr. 2013.
[10] Dat Tran, Wanli Ma, and Dharmendra Sharma, "Network Anomaly Detection using Fuzzy Gaussian Mixture Models", International Journal of Future Generation Communication and Networking, pp.37- 42, 2012.
[11] Vahid Golmah, "An Efficient Hybrid Intrusion Detection System based on C5.0 and SVM", International Journal of Database Theory and Application Vol.7, No.2, pp. 59-70, 2014.
[12] Punam Mulak, Nitin R. Talhar, "Novel Intrusion Detection System Using Hybrid Approach", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 4, No. 11, ISSN: 2277 128X, November 2014.
[13] Venkata Suneetha Takkellapati1, G.V.S.N.R.V Prasad, "Network Intrusion Detection system based on Feature Selection and Triangle area Support Vector Machine", International Journal of Engineering Trends and Technology, Vol. 3, No. 2012.
[14] Vaishali Kosamkar, Sangita S Chaudhari, "Improved Intrusion Detection System using C4.5Decision Tree and Support Vector Machine", International Journal of Computer Science and Information Technologies, Vol. 5, No. 2, pp. 1463- 1467, 2014.
[15] Harmeet Kaur, Ravneet Kaur, "Crossbreed Routing Protocol for SPEED Terminology in Wireless Sensor Networks", International Journal of Advance Research in Computer Science and management Studies, Vol. 2, No. 7, ISSN: 2321-7782, July 2014..
[16] H. Oh, I. Doh and K. Chae, "Attack classification based on data mining technique and its application for reliable medical sensor communication", International Journal of Computer Science and Applications, Vol. 6, No. 3, pp. 20-32, 2009.
[17] N. Ye and X. Li, "A Scalable Clustering Technique for Intrusion Signature Recognition", Proceedings of 2001 IEEE Workshop on Information Assurance and Security, 2001.

[18] G. Singh, F. Masseglia, C. Fiot, A. Marascu and P. Poncelet, "Data Mining for Intrusion Detection: from Outliers to True Intrusions", The 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'09), Thailand, 2009.
[19] K. Faraoun and A. Boukelif, "Genetic Programming Approach for Multi-Category Pattern Classification Applied to Network Intrusions Detection", The International Arab Journal of Information Technology, Vol. 4, No. 3, 2007.
[20] W. Lee, S. Stolfo, P. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop and J. Zhang, "Real Time Data Mining-based Intrusion Detection", Proceedings of DISCEX II, June 2001.
[21] K. Ioannis, T. Dimitriou and F. C. Freiling, "Towards Intrusion Detection in Wireless Sensor Networks", 13th European Wireless Conference, Paris, April 2007.
[22] D. Farid, J. Darmont, N. Harbi, N. Hoa and M. Rahman, "Adaptive
Network Intrusion Detection Learning: Attribute Selection and Classification", International Conference on Computer Systems Engineering (ICCSE 09), Bangkok, Thailand, December 2009.
[23] J. Zhang and M. Zulkernine, "Anomaly Based Network Intrusion Detection with Unsupervised Outlier Detection", Symposium on Network Security and Information Assuranc-Proc. of the IEEE International Conference on Communications (ICC), Istanbul, Turkey, June, 2006.
[24] M. Tavallaee, E. Bagheri, W. Lu and A. Ghorbani, "A Detailed Analysis of the KDD'99 CUP Data Set", The 2nd IEEE Symposium on Computational Intelligence Conference for Security and Defense Applications (CISDA), 2009.
[25] M. Campos and B. Milenova, "Creation and Deployment of Data Mining-Based Intrusion Detection Systems in Oracle Database 10g", an online document at http://www.oracle.com/technology/products/bi/odm/pdf/odm_based_intrusion_detection_paper_1205.pdf.
[26] Prothives and S. Srinoy, "Integrating ART and Rough Set Approach for Computer Security", Proceedings of the International MultiConference of Engineers and Computer Scientists, Vol. 1, 2009.
[27] H. Güneş Kayacık, A. Nur Zincir-Heywood and M. I. Heywood, "Selecting features for intrusion detection: a feature relevance analysis on KDD'99 intrusion detection datasets", Third Annual Conference on Privacy, Security and Trust, October 2005.
[28] M. Amini and R. Jalili, "Network-based intrusion detection using unsupervised adaptive resonance theory (ART)", Proceedings of the fourth conference on engineering of intelligent systems (EIS 2004), Madeira, Portugal, 2004.
[29] J. Xiao and H. Song, "A Novel Intrusion Detection Method Based on Adaptive Resonance Theory and Principal Component Analysis", Proceedings of the 2009 International Conference on Communications and Mobile Computing, Vol. 3, 2009.
[30] [30] Skoudis, Ed, and Tom Liston, "Counter hack reloaded: a step- by-step guide to computer attacks and effective defenses", Prentice Hall Press, 2005.
[31] K. Labib and V. Rao Vemuri, "Detecting Denial-of-Service And Network Probe Attacks Using Principal Component Analysis", In Third Conference on Security and Network Architectures, La Londe, (France), 2004.
[32] T. Eldos, M. Khubeb Siddiqui and A. Kanan "On the KDD'99 Dataset: Statistical Analysis for Feature Selection", Journal of Data Mining and Knowledge Discovery, 2012.