Artificial Intelligence Based Network Traffic Analysis to Handle Large-Scale and High Speed Traffic

Akshatha V[1]
[1] *PG Scholar,dept.of MCA*
Dayananda Sagar College of Engineering(VTU)
Bangalore,Karnataka,India-560078
mendonakshatha@gmail.com

Dr.Srinivasan V[2]
[2]*Assosiate Profeesor,dept.of MCA*
Dayananda Sagar College Of Engineering (VTU)
Bangalore, Karnataka,India-560078
srinivasan-mcavtu@dayanandasagar.edu

***Abstract*—** The network administrator is responsible for comprehensively analyzing network traffic and managing various applications utilized in the network. This includes tasks like network monitoring, anomaly detection, and optimizing network systems to extract valuable insights from the network traffic. It is important in core network for monitor the usage of network resources and also gives solution for problems. However when I retrieve the data from massive-scale network traffic it become challenging issue. Some software which does not support for retrieve information the network traffic manage the massive-scale network traffic such as NetFlow and Wire Shark are software are used for this problem. When I finding the solution for network traffic I practical experiments the methods for solution such as like many real-world datasets and the experimental results which functions which output will give extensive collection of innovative approach mainly which is used for large-scale information data. Detect the unnecessary activities and also Monitoring network traffic is main two task for managing by Computer Security Incident Response Teams (CSIRTs). CSIRTs is tool used for collect and monitor the network traffic data and also it focuses on the analyzing the data and also it detects the dangerous activities if it is happening means it will give procedure to solve that. This will give the effective way for manage the network traffic.

Keywords: Network Traffic Analysis, Text Retrieval Algorithm, And Information Retrieval.

I. INTRODUCTION

Nowadays, with help of development of communication technology, increasing of network bandwidth, this will lead to various complicated network security issues like malicious network information dissemination and privacy leakage. To address these issues, it is important to develop advanced solutions that can effectively manage network traffic and analyze it in a readable format, accommodating different languages and encoding formats[6]. While the specific software you mentioned in the last sentence is not clear, I can provide some general insights on how modern technologies and approaches are addressing these challenges. Enhanced Packet Analysis Tools: Network analysis tools are continuously evolving to handle large-scale and high-speed traffic[3]. They incorporate advanced algorithms and optimizations to process and analyze network packets more efficiently. These tools computing, and hardware acceleration to improve perfor Multilingual Support: To address the need for non-English languages, network analysis tools are incorporating multilingual support[2]. They can handle various encoding formats to accurately interpret and display content in different languages. This enables Empowering network administrators to examine network data irrespective of the language used. Deep Packet Inspection: Deep packet inspection (DPI) techniques are used to extract detailed information from network packets. DPI can analyze packet payloads, including text, images, and multimedia data, allowing for more comprehensive analysis[6]. With DPI, network administrators can search for specific keywords or patterns in the packet content, aiding in the identification of malicious network information. Artificial Intelligence and Machine Learning: AI and ML techniques are increasingly utilized in network security to enhance traffic retrieval and analysis[2]. These technologies can automate the identification and classification of network traffic, enabling faster detection of anomalies and potential security threats. Additionally, AI powered language processing models can aid in the interpretation of non-English content[9]. It's worth noting that the field of network security and analysis is constantly evolving, with new techniques and technologies emerging to address the increasing complexity of network traffic. By combining advanced packet analysis tools, multilingual support, deep packet inspection, AI/ML capabilities, and visualization techniques, network administrators can better manage network traffic and effectively mitigate security risks.

Various methods are available for network traffic analysis, including the software embedded mode, SNMP-based mode and hardware bypass mode[7]. Let's explore these methods from a different perspective, Software Embedded Mode: In this mode, traffic analysis software is installed on a host machine to capture and analyze network traffic. Examples of such software include Sniffer Pro and Wireshark, which intercept all packets on the host machine[3]. They can perform detailed analysis by parsing packet headers and content data. However, the content information is typically presented in hexadecimal or ASCII code[10]. This method is limited by the performance of the host machine and software design, making it suitable for small-scale and low-speed traffic analysis. SNMP-based Mode: SNMP based mode relies on Simple Network Management

Protocol (SNMP) and uses switches or network equipment to collect basic information about network traffic. Software like MRTG and NetFlow can provide statistical information, generating graphics and reports based on the collected data. While capable of control massive-scale traffic, this method lacks the ability to Retrieve network traffic content with restricted analysis depth Stream-based Mode: Similar to SNMP- based mode, stream-based mode collects information from switches or network equipment[1]. However, it focuses on analyzing traffic patterns and flows rather than individual packets.This method can handle large-scale traffic efficiently. However, like SNMP-based mode, it does not provide access to the content of network traffic. Hardware Bypass Mode: The hardware bypass mode entails capturing the entirety of the original network traffic by employing optical splitters or traffic mirroring replication devices, enabling comprehensive analysis of the traffic content[12]. Advancements in network applications have made it possible to obtain large-scale network traffic for in-depth analysis using this method. It offers flexibility in extracting specific information based on analysis goals[4]. However, the hardware cost associated with this mode is relatively high. In summary, these methods vary in terms of analysis performance, environment configurations, deep message analysis capabilities, and analysis flexibility. For a detailed comparison of these methods, please refer to Table I.

This paper introduces a system designed for retrieving and analyzing traffic data, incorporating multiple essential characteristics[5]. Notably, my system can convert unreadable Chinese content found in the original traffic into a readable format, such as text[7]. Additionally, I have developed an efficient retrieval algorithm that enables quick extraction of important information from the traffic, including text, IP addresses, and domain names[3]. In summary, In this paper research focuses on developing a system capable of deciphering and analyzing traffic data, making Chinese content understandable, and facilitating the retrieval of crucial information like text, IP addresses, and domain names.

Table 1:

	Performance	**Deep analysis**	**Cost**	**Flexibility**
Software embedded	Low	Strong	Low	Strong
SNMP	High	Weak	Low	Weak
Streambased	High	Weak	High	Weak
Hardware bypass	High	Strong	High	Strong

Fig 1: Network analysis and Retrieval Technology Comparison

This paper presents significant contributions in the following areas:

A. Development of LTARS: The authors have developed and implemented LTARS, a specialized system tailored for managing large-scale network traffic. LTARS includes features like protocol filtering, session reorganization, content extraction within sessions, and transformation into readable text format. Moreover, it facilitates efficient content retrieval from the network traffic.

Introduction of CFS text retrieval algorithm: The researchers introduce a new text retrieval algorithm named CFS, which demonstrates significant performance enhancements in the retrieval process within their system[3]. CFS is specifically designed to address the challenges of retrieving extensive textual information generated by massive-scale network traffic. In simpler terms, this paper introduces LTARS, a system capable of effectively managing and analyzing substantial amounts of network traffic. LTARS performs essential Performing functions like protocol filtering, session reorganization, content extraction, and transforming it into easily understandable text[12]. The paper also presents CFS, a new and efficient text retrieval algorithm that greatly enhances the system's ability to retrieve pertinent information from the extensive text data in network traffic[2]. The paper is organized as follows: Section II describes the LTARS system, followed by a comprehensive explanation of the CFS algorithm in Section III[3]. Section IV presents the performance evaluation and experimental results.

II. LITERATURE SURVEY

1) In this paper, a new algorithm was suggested for predicting network traffic by using a type of neural network called a Backpropagation (BP) neural network. Through simulations, it was observed that the proposed algorithm outperformed the conventional BP neural network in terms of prediction accuracy, yielding smaller errors.

2) The researchers conducted a comprehensive review and evaluation of multiple methods used for predicting network traffic. They thoroughly examined the unique characteristics and methodologies employed in previous studies. They also summarized the previous research conducted in the area of network traffic analysis and prediction. To accomplish this, they surveyed and studied earlier investigations focusing on network traffic analysis. They identified and discussed several approaches proposed for analyzing and predicting network traffic, including techniques such as data mining, neural networks, component analysis, as well as linear and nonlinear time series models.

3) In this research, researcher focused on understanding the analysis needs of large-scale network traffic and the existing techniques used for network traffic analysis. Based on their analysis, they proposed a system that aims to restore and retrieve network traffic data effectively. One important aspect of system is the development of an efficient retrieval algorithm. This algorithm is designed to retrieve relevant information from the network traffic data in an efficient and timely manner. They conducted experiments to evaluate

the effectiveness of proposed algorithm and validate its performance. Overall, their research aims to address the challenges associated with analyzing large-scale network traffic. By proposing a system for network traffic restoration and retrieval, and introducing an efficient retrieval algorithm, They believetheir work contributes to improving the analysis.

4) In this paper, the researchers' analysis of network traffic, utilizing an ANN model with the LM algorithm and time series analysis, highlighted the model's capability for accurate traffic prediction. This underscores its importance as an excellent and fundamental tool in managing internet traffic, empowering network administrators to make informed decisions and optimize network performance in real-time scenarios.

5) The researchers suggested a method to predict the packetloss rate (PLR) over time. This prediction is valuable formanaging network congestion effectively. They employed an artificial neural network as a predictive model and trained it using Particle Swarm Optimization (PSO) algorithm to ensure accurate PLR prediction. The researchers discovered that by accurately predicting thePLR, they could enhance the quality of real-time multimedia traffic and reduce congestion issues.

6) In this paper, the researchers used wavelet analysis and Hopfield neural network to create a model for predicting network traffic. The model was tested through simulations, which showed that it outperformed other methods in terms of accuracy. Moreover, the modelproved to be adaptable to different network situations, making it a valuable tool for forecasting future traffic patterns.

7) In this paper, the researchers discovered that when network traffic is left uncontrolled, it can lead to congestion and network paralysis. To tackle this issue, they used traffic forecasting technology to understand the changes in traffic patterns. They applied the ant colony algorithm to improve an existing prediction model called the gray model, resulting in the development of the IACGray algorithm. Their experiments confirmed that the improved IAC-Gray method provided more accurate predictions, making it avaluable tool for forecasting network traffic.

8) In this paper, the researchers discovered that traditional network traffic prediction models struggled to capture the complex and fluctuating nature of modern large- scale networks. To overcome this limitation, they proposed a new prediction model called MK-SVR. Experimental results confirmed that this model accurately described the changing trends in network traffic and significantly improved prediction accuracy by reducing errors. The MK-SVR model represents a valuable tool for predicting complex network trafficpatterns.

9) In simpler terms, the researchers introduced methods for detecting unusual situations in network traffic and evaluated their performance. They proposed a method that involved assessing the security level using a modified Exponential Moving Average and subjective logic opinions. This approach successfully identified and classified security-related problems in computer networks. The experimental results and statistical analysis demonstrated the effectiveness of employing Brown's exponential smoothing for forecasting, makingit a valuable tool for detecting abnormal situations in real world networks. Additionally, Brown's forecasting method was advantageous due to its lightweight nature, making it suitable for practical use in network analysis.

10) In simpler terms, the researchers developed an algorithmfor predicting small-scale network traffic. They used a local LSSVM regression model that was specifically tailored to the prediction task. By applying the Pattern Search method, they optimized the model's parameters. The algorithm selected a training set that was similar to the test set, filtering out irrelevant data. The researchers showcased the algorithm's effectiveness and efficiency, outperforming existing methods. They found that the prediction error was primarily concentrated close tozero.

III. PROPOSED METHODOLOGY

The proposed methodology for network traffic analysis using artificial intelligence refers to a systematic approach or framework that outlines the steps and techniques involved in leveraging artificial intelligence (AI) algorithms and techniques to analyze and understand network traffic data. It encompasses the entire process, from data collection and preprocessing to model selection, training, analysis, and prediction.

The primary objectives for network traffic analysis using artificial intelligence (AI) include:

1) Anomaly Detection: Detect and identify unusual or abnormal patterns in network traffic that could indicate security breaches, cyber attacks, or network performance issues. AI algorithms can learn from historical data and identify deviations from normal behavior, enabling proactive response and mitigation.

2) Traffic Classification: Classify network traffic into different categories based on its characteristics and behavior. This can help distinguish between different applications, protocols, or services running on the network, facilitating better network management, QoS (Quality of Service) optimization, and resource allocation.

3) Traffic Prediction: Forecast and predict future network traffic patterns, such as peak hours, traffic loads, or bandwidth demands. AI models can analyze historical data to identify trends and make accurate predictions, enabling capacity planning, network optimization, and efficient resourceprovisioning.

4) Real-time Monitoring and Alerting: Continuously monitor network traffic in real-time and generate alerts or notifications for critical events, abnormal behavior, or potential security incidents. AI-powered systems can

provide timely notifications, enabling prompt actions to address network issues and ensure network stability and security.

These objectives aim to leverage the power of AI in network traffic analysis to improve network performance, enhance security measures, optimize resource allocation, and enable proactive network management.

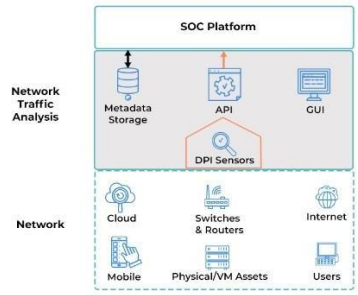

Fig 2. The overview of network traffic analysis

IV. ALGORITHM

An algorithm for network traffic analysis using artificial intelligence refers to a set of computational steps and procedures that leverage artificial intelligence techniques, such as Matching Multiple Patterns and the formation of the goto, output, and Skip constructs The algorithm aims to extract meaningful insights, patterns, or trends from the data, enabling various tasks such as traffic classification, anomaly detection, traffic prediction, or performance optimization.

A. The Algorithm for Matching Multiple Patterns

It also known as Algorithm 1, uses a goto function to map a state and an input character to another state. The purpose of this algorithm is to match patterns specified by the output function. When the algorithm encounters a mismatch between the input character and the current state, it uses the Skip function to select a new state and restarts the state transition process from that state. If the algorithm reaches a final state, it indicates Verifying the successful detection of the specified patterns by the output function. In simpler terms, the algorithm scans through a string and checks if it contains any predefinedpatterns. It keeps track of its current state and uses the goto function to determine the next state based on the current state and the input character. If there is a mismatch, the Skip function helps the algorithm choose a different state to continue the search for patterns. The algorithm stops when it finds a match for one of the patterns or reaches the end of the string.

Algorithm 1: Matching Multiple patterns.

Input: A text *str*[l : m] and *goto*, *output*, *Skip*.
Output: The locations where the matched patterns arefound within a provided string

```
x := min{ lengths of all patterns} ;
s := 0; while
x ≤m do
    if goto(s, str[i]) = 0 then
        x := x + Skip(str[x]);
    else
        s := goto(s, str[x]); if
        output(state) = NULL then
                print x;
    end x := x - 1;
    end
end
```

A proposed methodology for network traffic analysis using artificial intelligence typically involves several key steps:

A. Data Collection: Gather network traffic data from varioussources, such as network devices, sensors, logs, or packet captures. This data may include packet headers, payload information, flow records, or other relevant network metadata.

B. Data Preprocessing: Cleanse and preprocess the collecteddata to remove noise, handle missing values, and normalize the data. This step may involve techniques such as data filtering, feature extraction, and data transformation.

C. Analysis and Prediction: Apply the trained AI models to analyze network traffic patterns, detect anomalies, classify traffic types, predict future traffic behavior, or identify security threats. This step involves utilizing the trained models to make predictions or generate insights from the network traffic data.

D. Deployment and Monitoring: Deploy the trained AI models in a production environment to perform real-time network traffic analysis. Continuously monitor the performance and effectiveness of the deployed models, and update them as new data becomes available or network conditions change

By following this proposed methodology, network traffic analysis using artificial intelligence can help network administrators and security professionals gain valuable insights, enhance network management, improve security measures, andoptimize network performance.

V. RESULT AND DISCUSSION

Network traffic analysis using AI has delivered remarkable results by enhancing threat detection, enabling realtime anomaly detection, providing predictive insights, automating analysis processes, and offering scalability and adaptability. These outcomes empower organizations to strengthen network security, optimize performance, and make informed decisions toensure the efficient and secure operation of their networks.

Network traffic analysis faces several challenges and problems, including:

A. Data Complexity: Network traffic data is often complex and heterogeneous, consisting of various protocols, formats, and sources. Analyzing and interpreting this diverse data requires expertise and specialized techniques.

B. Security and Privacy: Network traffic analysis involves sensitive data, such as user information and communication content. Ensuring the security and privacy of this data during analysis poses challenges, particularlywhen dealing with encrypted traffic.

C. Traffic Anomalies: Detecting and understanding anomalous patterns in network traffic is crucial for normal and anomalous behavior can be challenging due to the dynamic nature of network traffic. Certain

applications require real-time network traffic analysis to respond promptly to emerging threats or performance issues. However, processing and analyzing traffic data in real-time can be demanding and resource intensive.

To address the challenges in network traffic analysis, severalsolutions and approaches can be implemented

A. Real-time Analysis Tools: Implement real-time analysis tools that can monitor and analyze network traffic in real-time. This allows for immediate detection of anomalies, performance issues, and security threats, enabling prompt response and mitigation.
B. Real-time Analysis Tools: Implement real-time analysis tools that can monitor and analyze network traffic in real-time. This allows for immediate detection of anomalies, performance issues, and security threats, enabling prompt response and mitigation.
C. Privacy and Security Considerations: Implement appropriate measures to ensure the privacy and securityof network traffic data during analysis. This may involve encryption, anonymization techniques, and adherence to data protection regulations
D. Collaboration and Knowledge Sharing: Foster collaboration and knowledge sharing among network administrators, analysts, and researchers. Sharing best practices, insights, and threat intelligence can help improve the overall network traffic analysis capabilities.

In simpler terms, I am examining the needs for analyzing large amounts of network traffic and the techniques currently used for analyzing such traffic. I then suggest a system that can restore and retrieve network traffic data. I investigate an effective algorithm for retrieving information efficiently. I provide experimental results to demonstrate and confirm its effectiveness.

VI. REFERENCE

[1] Ming Zhang and Yanhong Lu, (2015), " Adaptive Network Traffic Prediction Algorithm based on BP Neural Network", International Journal of Future Generation Communication and Networking Vol. 8, No.5 (2015), pp. 195-206.

[2] Manish R. Joshi et. Al.,(2012) "A Review of Network Traffic Analysis and Prediction Techniques", pp. 1-22.

[3] Ting Han1,3, Yuanming Zhang1, Hezhen Li1, Xiaoyu Zhang1, Jing Tao , The study on Large-scale Network Traffic Analysis and Retrieval System Using CFS Algorithm 2019 IEEE Intl Conf on Dependable.

[4] Samira Chabaa, Abdelouhab Zeroual, Jilali Antari, (2010), " Identification and Prediction of InternetTraffic Using Artificial Neural Networks", J.Intelligent Learning Systems & Applications, 2010, 2,147-155.

[5] Manish, P. Ganvir, Dr. S.S.Salankar, (2015), "Time Series Forecasting of Packet Loss Rate Using Artificial Neural and General Science Volume 3, Issue 2, pp. 466-472.

[6] Han Song, Luying Gan (2015), "The Research on the Prediction of the Network Traffic Based on the Improved IAC-Gray Method", CHEMICAL ENGINEERING TRANSACTIONS, VOL. 46, pp. 1297-132.

[7] Han Song, Luying Gan (2015), "The Research on the Prediction of the Network Traffic Based on the Improved IAC-Gray Method", CHEMICAL ENGINEERING TRANSACTIONS, VOL. 46, pp. 1297-132.

[8] Sun Guang, (2013), "Network Traffic Prediction Basedon the Wavelet Analysis and Hopfield Neural Network", International Journal of Future Computer and Communication, Vol. 2, No. 2, April 2013.

[9] Changsheng Xiang; Peixin Qu, Xilong Qu, (2015), "Network Traffic Prediction Based on MKSVR", Journal of Information & Computational Science 12:8(2015) 3185–3197

[10] Jarosław Bernacki, et.al.,(2015), " Anomaly Detection in Network Traffic Using Selected Methods of Time Series Analysis", J. Computer Network and Information Security, 2015, 9, 10-18.

[11] Tao Peng and Zhoujin Tang,(2015), "A Small Scale Forecasting Algorithm for Network Traffic based on Relevant Local Least Squares Support Vector Machine Regression Model", Appl. Math. Inf. Sci. 9, No. 2L, 653-659.

[12] Huan Luo* , Tiankui Zhang*, Yong Sun*, Chunyan Feng*, and Weidong Fengt , As study on Two Dimensional Cooperation Prediction Algorithm of Communication Network Traffic in Smart Grid 2015 10th International Conference on Communications and Networking in China (China Com)

[13] Yun Lan, Yong Sun, Sheng-peng Liu, Zhong-zheng Ma, As study on A Real-Time Network Traffic Analysis and QoS Management Platform 2017 9th IEEE International Conference on Communication Software and Networks. andCommunication, Vol. 2, No. 2, April 2013.