Multilayer Seasonal Autoregressive Integrated Moving Average Models for Complex Network Traffic Analysis

[1]Prathipa Ravanappan, [2]Maragatharajan M, [3]Rashika Tiwari, [4]Srihari T and [5]Lavanya K
[1]Department of Electronics and Communication Engineering, Panimalar Engineering College, Chennai, India.
[2]School of Computing Science and Engineering (SCSE), VIT Bhopal University, Madhya Pradesh, India.
[3]Geetanjali Olympiad School, Vartur Hobli, Bengaluru, India.
[4]Department of Electrical and Electronic Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (SIMATS), Saveetha University, Chennai, Tamil Nadu, India
[5]Department of Electronics and Communication Engineering, Velammal Engineering College, Chennai, India.
[1]rprathipa@ieee.org, [2]maragatharajanm@gmail.com, [3]rashika.tiwari17@gmail.com,
[4]k.t.srihari@gmail.com, [5]lavanya201180@gmail.com.

Correspondence should be addressed to Prathipa Ravanappan : rprathipa@ieee.org.

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404023
Received 16 April 2023; Revised from 19 July 2023; Accepted 06 December 2023.
Available online 05 January 2024.

Abstract – The ever-increasing amount of network traffic generated by various devices and applications has made it crucial to have efficient methods for analyzing and managing network traffic. Traditional approaches, such as statistical modeling, have yet to be proven enough due to network traffic's complex nature and dynamic characteristics. Recent research has shown the effectiveness of complex network analysis techniques for understanding network traffic patterns. This paper proposes multilayer seasonal autoregressive integrated moving average models for analyzing and predicting network traffic. This approach considers the seasonal patterns and interdependencies between different layers of network traffic, allowing for a more accurate and comprehensive representation of the data. The Multilayer Seasonal Autoregressive Integrated Moving Average (MSARIMA) model consists of multiple layers, each representing a different aspect of network traffic, such as time of day, day of week, or type of traffic. Each layer is modeled separately using SARIMA, a popular time series forecasting technique. The models for different layers are combined to capture the overall behavior of network traffic. The proposed approach has several benefits over traditional statistical approaches. It can capture network traffic's complex and dynamic nature, including short-term and long-term seasonal patterns. It also allows for the detection of anomalies and the prediction of future traffic patterns with high accuracy.

Keywords – Network, Traffic, Statistical Modeling, Dynamic Characteristics, Seasonal Patterns.

I. INTRODUCTION

The internet and its additional technologies have significantly expanded our communication ability and access information. As a result, our world has become increasingly interconnected, and the amount of data and information transmitted on the network has reached unprecedented levels. With the growth of data, there has also been a corresponding increase in network traffic [1]. Network traffic refers to the data flow between devices connected to a network, such as computers, servers, and other devices. It encompasses all the data that is exchanged between these devices, including emails, file transfers, web browsing, streaming media, and more [2]. Network traffic is continuously increasing, and as our dependence on the internet grows, it has become essential to understand its complexities and demands better. Complex network traffic analysis is studying and understanding the patterns and characteristics of data flow across a network [3]. It involves collecting, analyzing, and interpreting large amounts of data to gain insight into how a network is utilized. This analysis helps identify and address network issues, improve performance, and optimize network resources [4]. One of the primary goals of complex network traffic analysis is to identify and manage network congestion. As network traffic increases, there is a higher chance of congestion, leading to delays in data transmission [5].

In some cases, it can lead to network failures and disrupt communication. Network traffic analysis helps identify areas of congestion and allows network administrators to take necessary measures to alleviate it [6]. It could involve rerouting traffic or implementing quality-of-service measures to prioritize critical data. Another crucial aspect of complex network traffic analysis is identifying potential security threats. With the increasing flow of data, there is a higher risk of cyber-attacks and network breaches [7]. Network traffic analysis can help identify unusual or suspicious data patterns that could indicate malicious activity. This information can be used to strengthen network security and prevent potential threats. Complex network traffic analysis also plays a crucial role in optimizing performance and resource allocation [8]. Understanding and identifying traffic patterns and data-intensive processes enable network administrators to allocate resources more effectively. It helps improve network speed and reduce the risk of network failures.

Network traffic analysis can also provide valuable insights for businesses. Organizations can identify usage trends and patterns by understanding the data flow across a network, which can inform decision-making processes [9]. This knowledge can help optimize network usage, reduce costs, and improve productivity. In the modern age, where the internet has become an indispensable part of our daily lives, complex network traffic analysis has become more critical than ever. As the volume and complexity of network traffic grows, it is essential to understand and manage it effectively [10]. Not only does it help improve network performance and security, but it also provides valuable insights that can drive business growth and success. By continuously monitoring and analyzing network traffic, we can ensure a fast, secure, and efficient flow of data, enabling us to take full advantage of the internet's vast resources.

The main contribution of the research has the following,

- Identification of network abnormalities: Complex network traffic analysis helps identify any unusual or abnormal traffic patterns in a network. It can be caused by malicious activities such as cyber-attacks or network failures. By identifying these abnormalities, it is possible to take immediate action to mitigate potential risks or failures and ensure the smooth functioning of the network.
- Performance optimization: By analyzing the traffic flow and patterns, it is possible to optimize the performance of a network. It involves identifying bottlenecks and congestion points and implementing appropriate measures to improve network speed and efficiency.
- Resource allocation: Analysis of network traffic can help identify the usage of network resources such as bandwidth, storage, and processing power. This information can be used to allocate resources more efficiently and reduce the likelihood of resource overutilization or wastage.
- Prediction of network usage: Complex network traffic analysis can also provide insights into a network's usage trends and patterns. It can help predict network capacity requirements and planning.

II. MATERIALS AND METHODS

Complex Network Traffic Analysis is an emerging field that combines the study of computer networks and traffic patterns to understand and manage network behavior. However, analyzing network traffic has become increasingly challenging with the increasing use of advanced technologies [11]. There are several issues when analyzing complex network traffic, which are discussed below. One of the main issues in complex network traffic analysis is the sheer volume of data generated. As technology has advanced, so has the amount of traffic that networks can handle. It makes it challenging to analyze and make sense of the data generated, which can be in terabytes or even petabytes [12]. This high volume of data poses challenges for researchers regarding storage capacity, computational power, and processing time. The network traffic is not a regular data flow and contains a mix of different data types, including text, images, audio, and video [13]. Analyzing this diverse and heterogeneous data requires specialized tools and techniques to handle different file formats and data types. It also adds to the complexity of the analysis process [14].

An issue in complex network traffic analysis is the real-time nature of network traffic. Network traffic constantly changes, and patterns can emerge and disappear within seconds. Therefore, it is crucial to have real-time monitoring and analysis systems in place to detect and respond promptly to any anomalies or malicious activities [15]. Furthermore, network traffic is only sometimes generated by legitimate users and devices. With the rise of cyber threats and attacks, there has been an increase in malicious traffic that can significantly impact network performance and security [16]. Detecting and filtering out this malicious traffic is a significant challenge for researchers as it requires sophisticated algorithms and techniques to differentiate between normal and abnormal network behavior. Another challenge is the need for more standardization in network traffic data. Different network devices and protocols generate data in different formats, making integrating and analyzing the data effectively challenging [17]. This lack of standardization can also hinder the interoperability of different network analysis tools and systems, making it difficult to have a unified approach to network traffic analysis. In addition to these technical issues, ethical and privacy concerns need to be addressed when analyzing network traffic data [18]. As network traffic data contains sensitive information, it is necessary to have proper protocols and regulations in place to protect user privacy and data confidentiality. The complex network traffic analysis field is facing several challenges due to the complexity and diverse nature of network traffic data [19]. It must develop advanced tools and techniques to handle the high volume and heterogeneous data to make sense of the ever-changing network traffic patterns successfully. Moreover, addressing the ethical and privacy concerns associated with network traffic analysis is essential to maintain the balance between network security and user privacy [20]. Here some of the performance issues were identified. They are,

- Network Congestion: Complex networks experience high traffic volumes, leading to congestion and slowdowns in data transmission.
- Bottlenecks: Due to the interconnected nature of complex networks, traffic bottlenecks can occur at specific nodes, resulting in delays and reduced performance.
- Network Security Threats: Complex networks are vulnerable to various security threats, such as malware, hacking, and DDoS attacks, which can compromise data integrity and disrupt network services.
- Protocol Incompatibilities: Different devices and protocols may not be fully compatible in heterogeneous complex networks, resulting in communication issues and network failures.
- Quality of Service Issues: With diverse applications and services competing for network resources, It can be adversely affected, leading to poor performance and user dissatisfaction.
- Network Monitoring and Management Challenges: Due to the scale and complexity of network traffic, it can be challenging to monitor and manage effectively, making it difficult to identify and address issues promptly.

One of the main novel aspects of complex network traffic analysis using multilayer seasonal autoregressive integrated moving average models is its ability to simultaneously analyze network traffic data at multiple layers. Traditional methods for network traffic analysis typically focus on one layer at a time, such as the application layer or transport layer. However, with modern networks' increasing complexity and diversity, it has become crucial to analyze traffic data at multiple layers to gain a more comprehensive understanding of network behavior. The use of MSARIMA models for complex network traffic analysis is a novel and innovative approach that can significantly improve our understanding of network behavior and enhance network performance and security. It represents a significant step forward in the field of network traffic analysis. It can benefit various industries and sectors that rely on complex networks for their operations significantly.

III. PROPOSED MODEL

MSARIMA models provide a powerful tool for this type of analysis, as they can capture the temporal and seasonal patterns of network traffic data at different layers. It allows for a more nuanced and accurate analysis of complex network dynamics. These models can help identify hidden relationships and dependencies between different layers and their impact on overall network performance. The application of MSARIMA models in network traffic analysis offers a unique approach to monitor and detect anomalies or abnormalities in network traffic data. By considering multiple layers of traffic data, the models can detect abnormal patterns or trends that may not be visible when analyzing only a single layer. It can significantly enhance the security and management of networks, as it allows for early detection and mitigation of potential issues. The proposed block diagram has shown in the following **Fig 1.**

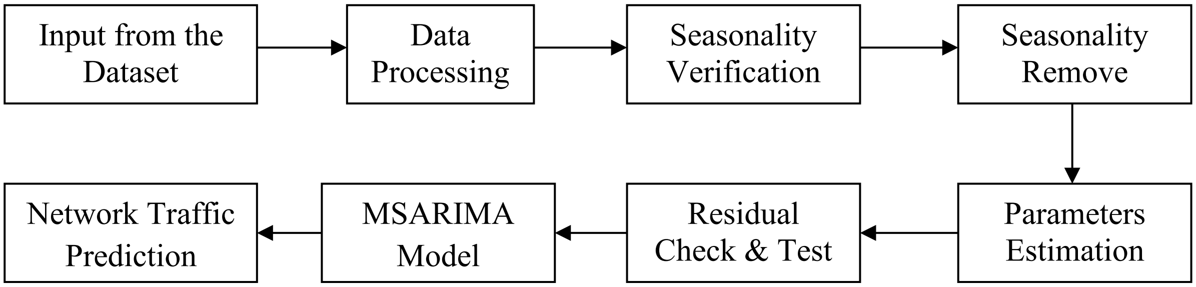

Fig 1. Proposed Block Diagram

Input from the dataset is the first step in network traffic prediction. This data is collected over time and includes information such as network traffic volume, usage patterns, and other relevant data points. This data is then used for data processing, which involves cleaning and organizing the data to make it usable for analysis. Seasonality verification is an essential step in network traffic prediction as it helps identify patterns in the data that occur at regular intervals, such as daily, weekly, or monthly. It is essential because network traffic often exhibits seasonality, following a recurring pattern. By verifying and understanding the seasonality in the data, it becomes easier to identify and remove it from the dataset. Once the seasonality has been identified and removed from the dataset, the next step is parameter estimation. It involves using statistical methods to estimate the data parameters, such as the mean and variance, which helps to characterize the dataset and make predictions based on the data patterns. Residual checks and tests are important steps in the process as they help evaluate the model's accuracy. It involves checking the difference between the predicted and actual values to determine the model's effectiveness in predicting network traffic. MSARIMA model is a commonly used statistical model for network traffic prediction. It considers seasonality, trends, and other essential factors to make accurate predictions.

This model combines the components of the Autoregressive Integrated Moving Average model with the ability to handle multiple time series data. Network traffic prediction is a continuous process, as the accuracy of the predictions can be improved by continuously re-evaluating and refining the model based on new data. This continuous improvement process is made possible by using continuous functions such as input from the dataset, data processing, seasonality verification, seasonality removal, parameter estimation, and residual check and test. The model can adapt to changing network traffic patterns and provide more accurate predictions by repeating these steps. This continuous process of network traffic prediction helps businesses and organizations make informed decisions about their network infrastructure and improve overall network performance.

Data Processing

Data processing plays a crucial role in complex network traffic analysis (CNTA) by facilitating large, heterogeneous dataset’s collection, organization, and analysis. The primary function of data processing in CNTA is to convert raw network data into a suitable format for analysis. It includes filtering out irrelevant or redundant data, cleaning up the data by removing errors or outliers and transforming the data into a standard format for consistency.

$$\frac{\partial O}{\partial P} * \frac{\partial P}{\partial O} = \left\{\frac{\partial}{\partial P}(e^{O} * P\, cos\, O_p)\right\} * \left\{\frac{\partial}{\partial O}(e^{P} * P\, sin\, O_p)\right\} \tag{1}$$

$$\frac{\partial O}{\partial P} * \frac{\partial P}{\partial O} = 1 \tag{2}$$

Another critical function of data processing in CNTA is data aggregation. It combines multiple data streams from different sources, such as network logs, network packets, or network flow data, into a single dataset.

$$\frac{\partial p}{\partial o} = \left(O * \frac{\partial P}{\partial o}\right) + \left(N * \frac{\partial O}{\partial p}\right) \tag{3}$$

Aggregation is necessary to gain a comprehensive view of network traffic and identify impossible patterns and anomalies with individual data sources. Data processing also involves data integration, combining data from different sources.

$$\frac{\partial p}{\partial o} = \left(e^{o} * \frac{\partial}{\partial o} cos\, O\, p\right) + \left(cos\, O\, p * \frac{\partial}{\partial o}(e^{o})\right) \tag{4}$$

$$\frac{\partial p}{\partial o} = (O * e^{o}\, sin\, O\, p) + (e^{o}\, cos\, O\, p) \tag{5}$$

In CNTA, this can include integrating data from different network protocols, devices, or systems. It allows analysts to understand the network holistically and identify relationships and dependencies between different components.

$$O = e(p) = p^{o} \tag{6}$$

Enriched data can provide valuable context to network traffic, enabling analysts to understand and interpret the data better.

$$\partial p'' = \lim_{p \to 0} \left(\frac{\partial p(o+p) - \partial p(o)}{\partial o}\right) \tag{7}$$

Data processing also plays a crucial role in data reduction and compression. It involves reducing the size of large datasets by summarizing or aggregating the data without losing important information.

$$\partial p' = \lim_{p \to 0} \left(\frac{\partial o^{p+o} - \partial p^{o}}{\partial o}\right) \tag{8}$$

It is essential in CNTA due to the immense volume of data in complex networks. Once the data has been processed and organized, the final function of data processing in CNTA is analysis.

$$\partial p'' = \lim_{p\to 0}\left(\frac{\partial(p^o * p^o) - \partial p^o}{\partial o}\right) \tag{9}$$

$$\partial p = \lim_{o\to 0}\left(\frac{\partial p^o * \partial(p^o - 1)}{\partial o}\right) \tag{10}$$

It can involve performing statistical analysis, machine learning algorithms, or visualization techniques to uncover insights and detect patterns in network traffic.

$$\partial o'' = \partial p^o * \lim_{o\to 0}\left(\frac{\partial(p^o - 1)}{\partial p}\right) \tag{11}$$

$$\partial o = \partial p^o * ln(p) \tag{12}$$

The analysis output can then be used for network troubleshooting, threat detection, and network performance optimization tasks. In addition to collecting and organizing data, data processing in CNTA also includes data enrichment. It involves enhancing the raw data with additional information such as geo-location, device type, or user information.

Seasonality Verification

Seasonality verification is an essential aspect of complex network traffic analysis that helps understand patterns and trends in network data. It involves identifying and analyzing patterns that occur at regular intervals, also known as seasonal patterns, in network traffic data.

$$\left(\frac{\partial O * \partial O_o}{\partial P_o}\right) = \frac{1}{2}\partial O * \partial p_o^2 \tag{13}$$

It can be done using statistical methods, such as time series analysis, to identify long-term trends and patterns in the traffic data. One of seasonality verification's primary functions is accurately predicting network traffic patterns over time.

$$\partial p_o^2 = \left(\frac{\partial O * \partial O_o}{\partial P_o}\right) * \frac{2}{\partial o} \tag{14}$$

Understanding the seasonal patterns in network traffic data, network administrators and analysts can anticipate and plan for changes in traffic volume and patterns. This information can be valuable in ensuring the network is configured correctly and can handle high traffic levels during peak seasons. Seasonality verification can help to identify anomalies or irregularities in network traffic data. Analyzing seasonal patterns can quickly identify any abnormal or unexpected changes in network traffic. It can detect and mitigate potential security threats or network failures.

Seasonality Remove

Seasonality removal is a data preprocessing technique that removes periodic patterns from time series data. It is a critical step in complex network traffic analysis as it helps to identify meaningful patterns and trends in network traffic data. Seasonality refers to regular fluctuations in the data, such as daily or weekly patterns, that can obscure the underlying behaviors and anomalies in network traffic. By removing seasonality, analysts can focus on the non-seasonal components of the data and better understand the true nature of network traffic. One of the critical functions of seasonality removal in complex network traffic analysis is to improve the accuracy of traffic forecasting. Seasonal patterns can make it challenging to predict future network traffic trends accurately.

$$\partial p_o^2 = \left(\frac{2 * \partial O_p}{\partial P_o}\right) \tag{15}$$

$$where, o = \left(\frac{\partial P_o}{\partial O_p^2}\right); \tag{16}$$

Removing these patterns, analysts can better identify long-term trends and make more accurate predictions about future traffic patterns. It is essential for network planning and optimization, where accurate forecasting is crucial for ensuring efficient network performance and resource allocation. Another function of seasonality removal is identifying anomalies or abnormal patterns in network traffic.

$$\partial o_p^2 = 2 * \partial o * \partial O_p \tag{17}$$

$$\partial o_p = \sqrt{2 * \partial o * \partial P_o} \tag{18}$$

Anomalies can indicate network failures, cyber-attacks, or other unusual events that require immediate attention. However, these anomalies can be masked by seasonal patterns in the data. By removing seasonality, analysts can more easily detect and investigate these anomalies and take appropriate actions to mitigate potential risks to the network. Seasonality removal also helps to simplify the analysis of complex network data. Removing seasonal patterns makes the data more accessible to interpret and visualize, making it easier to uncover essential insights and relationships. It is precious in complex networks where large amounts of data are collected and analyzed. By eliminating unnecessary noise and focusing on the non-seasonal components, analysts can better understand the underlying behaviors and dynamics of the network traffic. Seasonality removal is crucial in complex network traffic analysis by improving forecasting accuracy, identifying anomalies, and simplifying data analysis. It allows a better understanding of network traffic patterns and helps analysts make informed decisions to optimize network performance and security.

Parameters Estimation

Parameter estimation is an essential tool for analyzing complex network traffic. It involves determining the characteristics of a network, such as its structure, flow, and behavior, by measuring and analyzing various parameters. This process helps understand how the network operates, identify potential issues or bottlenecks, and optimize its performance. One of the main functions of parameter estimation is to capture the structure of a network. It includes identifying the different nodes and their connections, as well as the overall topology of the network. By analyzing parameters such as degree distribution, centrality measures, and clustering coefficients, researchers can gain insights into the organization and layout of a network. This information is crucial in understanding how information flows within the network and the potential pathways for communication. Another essential function of parameter estimation is measuring traffic flow within the network. It involves analyzing traffic volume, packet size, and distribution parameters. By doing so, researchers can identify the most heavily used paths and nodes, potential congestion points, and areas where the network may struggle to handle the traffic load. This information can help optimize the network infrastructure and identify potential security vulnerabilities.

Residual Check

Residual Check is a function used to analyze complex network traffic that helps identify anomalies or unusual patterns in the data. It compares the expected values of a particular network metric with the actual observed values. If there is a significant difference between the two, it indicates that there may be some anomalous behavior in the network. The first step in performing a residual check is establishing a baseline for the network. It involves collecting and analyzing historical data to determine the typical patterns and behaviors of the network. Once this baseline is established, the residual check function can monitor the network in real-time and identify any deviations from the expected patterns. The function works by calculating the difference between the current value of a network metric and the expected value based on the established baseline. The more significant the difference, the more likely there is a problem or anomaly in the network. For example, a sudden increase in network traffic may indicate a potential denial of service (DoS) attack, while a sudden decrease could indicate a system failure. Residual check is a critical aspect of network traffic analysis as it allows network administrators to quickly identify and respond to any unusual activity in the network. Detecting anomalies in real time can identify and mitigate potential threats before they cause significant damage to the system.

Test for Complex Network Traffic Analysis

The Test for Complex Network Traffic Analysis is another essential function for analyzing network traffic. It involves running tests on the data to assess its integrity, accuracy, and completeness. These tests can include statistical analysis, graph theory, and machine learning techniques. The primary purpose of the test is to ensure that the data being analyzed is reliable and can be used to make informed decisions about the network. It helps to identify any errors or inconsistencies in the data that could affect the accuracy of the analysis results.

$$p^{''}(o) = \lim_{o \to 0} \left(\frac{p(p+o) - p(o)}{o} \right) \tag{19}$$

One of the key advantages of using the Test for Complex Network Traffic Analysis is that it can help to identify hidden relationships and patterns in the data.

$$p^{'}(o) = \lim_{p \to 0} \left(\frac{p^{p+o} - p^{p}}{o} \right) \tag{20}$$

These patterns may not be apparent upon initial inspection. However, with the help of advanced analytical techniques, they can be uncovered and used to gain a deeper understanding of the network.

$$p(o) = \lim_{p \to 0} \left(\frac{(p^p * p^o) - p^p}{o} \right) \tag{21}$$

In addition to ensuring data integrity, the Test for Complex Network Traffic Analysis also helps identify areas where the network could be improved. For example, it may reveal bottlenecks or inefficiencies in the network that could be causing performance issues.

$$p(o) = e^p * \lim_{p \to 0} \left(\frac{1 - e^p}{o} \right) \tag{22}$$

Residual Checks and Tests for Complex Network Traffic Analysis functions are vital in ensuring the security and efficiency of complex networks. They help detect anomalies and assess data reliability, providing network administrators with the necessary information to make informed decisions about network management and optimization.

MSARIMA Model

The MSARIMA model is a powerful tool for analyzing complex network traffic, which refers to the flow of data or information between two or more connected devices or systems. This model combines the strengths of both ARIMA and SARIMA models to capture the traffic data's temporal dynamics and seasonal patterns. One of the critical functions of the MSARIMA model is its ability to capture the autocorrelation present in network traffic data. In other words, it considers that the previous traffic flow influences the current traffic flow. The proposed flow diagram has shown in the following **Fig 2.**

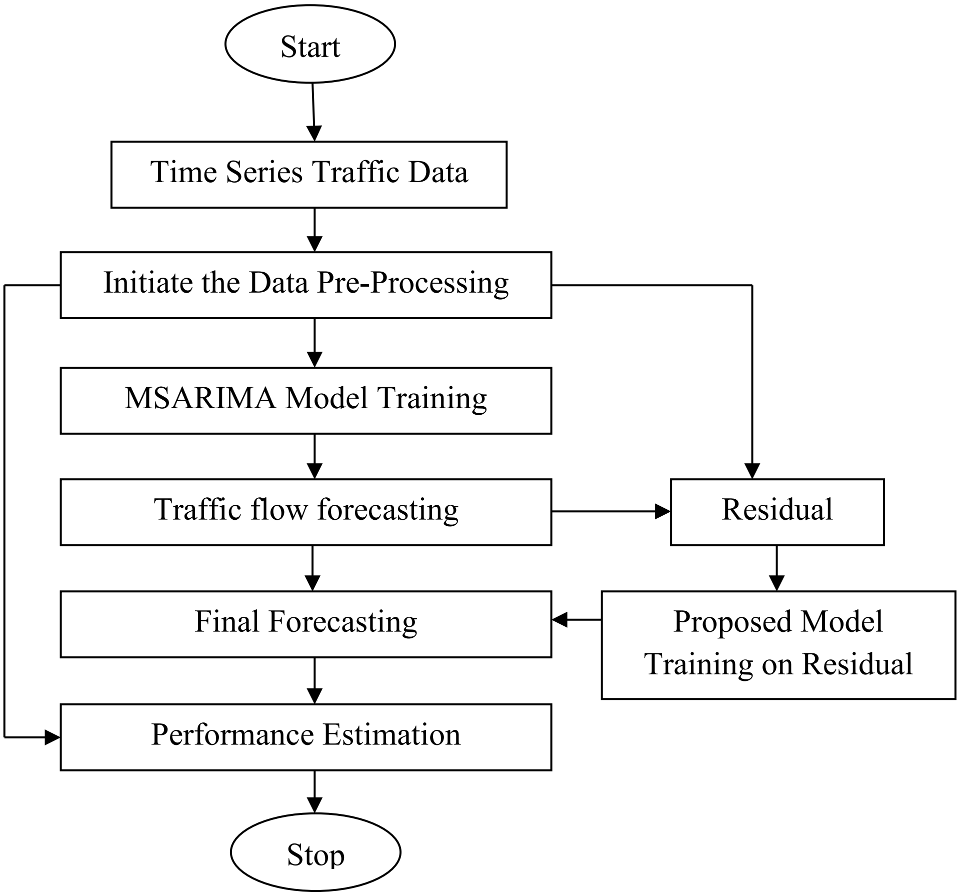

Fig 2. Proposed Flow Diagram

Various factors, such as network congestion, user behavior, or network events, can cause this autocorrelation. By incorporating this into the model, the MSARIMA can provide more accurate and reliable predictions of future traffic patterns. Another essential function of the MSARIMA model is its ability to handle seasonal patterns in network traffic. Many network systems experience periodic changes in traffic flow, such as daily or weekly spikes, which can be challenging to capture with traditional time series models. The seasonal component of the MSARIMA model allows it to model and predict these patterns effectively, helping network administrators better manage and plan for peak traffic periods.

The model also considers the impact of external factors such as holidays, events, or special promotions on network traffic. It makes it a valuable tool for analyzing and predicting traffic patterns during these periods, which can be crucial for businesses or organizations relying on their network systems. One of the unique features of the MSARIMA model is its ability to handle multiple input variables. It means that the model can incorporate other relevant information, such as weather conditions, economic indicators, or network capacity, in addition to the network traffic data. By considering these additional factors, the model can provide more accurate and comprehensive insights into the complex dynamics of network traffic.

Network Traffic Prediction

Traffic prediction is a vital function of complex network traffic analysis. It involves using machine learning and statistical techniques to forecast the future behavior of network traffic. This function can provide valuable insights and help manage the network efficiently. Firstly, traffic prediction helps in resource allocation. By analyzing the historical data and predicting future traffic patterns, network administrators can allocate resources such as bandwidth, servers, and routers accordingly. It ensures the network can handle the expected traffic without disruptions or delays. It also helps avoid underutilization of resources, leading to unnecessary expenses. Secondly, it aids in network optimization. By accurately predicting future traffic, network administrators can identify potential network congestion points and take proactive measures to alleviate them. It can include adding extra bandwidth, rerouting traffic, or implementing QoS policies.

These actions can improve network performance, reduce latency, and ensure a seamless user experience. Thirdly, traffic prediction can play a crucial role in network security. Analyzing traffic patterns and predicting anomalies can help in the early detection and prevention of malicious activities such as DDoS attacks, data breaches, and unauthorized access. This proactive approach can save valuable time and resources in handling potential security threats. Traffic prediction is vital for capacity planning. By forecasting future traffic, network administrators can plan for potential growth and plan for scaling up the network infrastructure accordingly. It ensures that the network can accommodate increasing traffic demands without any downtime. Traffic prediction is crucial for decision-making. Network administrators can make informed decisions about network upgrades, expansions, and investments by understanding past and current traffic patterns. It also helps identify areas for improvement and optimize network performance.

IV. RESULTS AND DISCUSSION

Traffic prediction is a crucial function of complex network traffic analysis. It helps in resource allocation, network optimization, security, capacity planning, and decision-making. By accurately predicting future traffic, it enables efficient management of the network and ensures a seamless user experience. The proposed model has compared with the existing Deep learning model under complex network (DLMCN), complex network analysis for traffic prediction (CNATP), complex network approach (CNA) and Chaotic characteristic analysis (CCA). The network traffic dataset [21] has been used for simulation purposes. Here, 80% data is used for training purpose and 20% data used for the testing purpose. The network simulation (NS-3) has been used for the simulation tool. **Table 1** shows the simulation parameters.

Table 1. Simulation Parameters

Parameters	Values
No. of Traffic Source	20
No. of resource blocks	12
Transmission data rate	120Mbps
Link capacity	12 Mbps
Link delay	25 ms
Transmission Protocol	TCP
Simulation Duration	100 s

Computation of Accuracy

Complex network traffic prediction is forecasting traffic trends and patterns in a network using machine learning algorithms and statistical analysis. The accuracy of such predictions is essential in ensuring efficient network management and resource allocation, which in turn can improve user experience and reduce network downtime. **Fig 3** shows the comparison of accuracy.

The computation of accuracy for complex network traffic prediction involves comparing the forecasted values with the actual values of network traffic. The most used metric for accuracy is the Mean Absolute Percentage Error. This metric measures the percentage difference between the predicted and actual values, making it a reliable measure of how well the predictions match the observed data. To calculate the MAPE for complex network traffic prediction, the forecasted values for a specific period are subtracted from the actual values, and the resulting absolute differences are divided by the actual values. This value is then multiplied by 100 to get a percentage. The MAPE is calculated for each data point in the prediction set and then averaged to get an overall accuracy score.

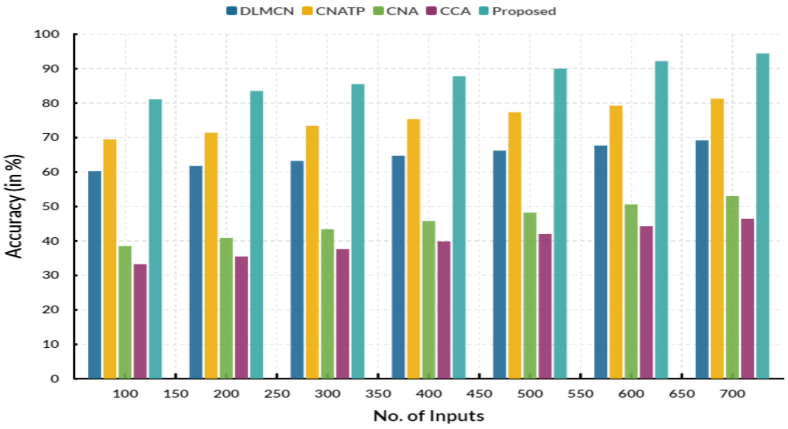

Fig 3. Comparison of Accuracy

Computation of Precision

Precision is a metric used to evaluate the accuracy of predictions in complex network traffic. It measures the proportion of correctly predicted traffic out of all predicted traffic for a specific class or category. To compute precision, we first need to divide the predicted traffic into different classes or categories based on specific characteristics such as source or destination, type of traffic, or time of occurrence. Then, for each class, we count the number of correctly predicted traffic instances and divide it by the total number of predicted instances. **Fig 4** shows the comparison of precision.

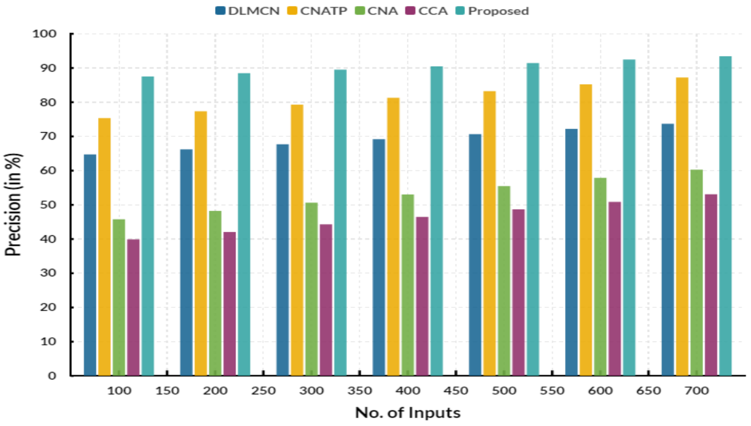

Fig 4. Comparison of Precision

It gives us the precision value for that specific class. The overall precision for the network traffic prediction is then calculated by averaging the precision values for all classes. A higher precision value indicates a more accurate prediction, while a lower value implies more incorrect predictions. This metric is essential in evaluating the performance and effectiveness of prediction models in handling complex network traffic. It can better predict and manage network traffic by improving precision, leading to more efficient and reliable network operations.

Computation of Recall

Recall is a metric used to evaluate the performance of a prediction model, specifically in the context of binary classification. In the case of Complex Network Traffic prediction, this metric measures the ability of the model to correctly identify all the relevant network traffic events. To compute recall, we need to understand the concept of true positives (TP), which refers to the cases where the model predicted a positive outcome, and the actual outcome was also positive. **Fig 5** shows the comparison of recall.

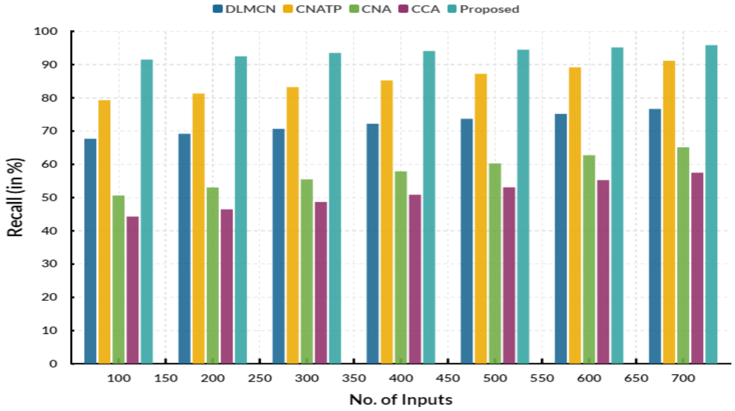

Fig 5. Comparison of Recall

In the context of Complex Network Traffic, this would mean predicting the occurrence of a relevant network event and that the event did indeed occur. On the other hand, false negatives (FN) refer to cases where the model predicted a negative outcome, but the actual outcome was positive. In the case of Complex Network Traffic, this would mean the model failed to predict a relevant event that occurred.

Computation of Traffic Discovery Rate

The traffic discovery rate measures how effectively a complex network traffic prediction model can identify and predict traffic patterns in a network. It is a crucial metric for evaluating the performance of a traffic prediction model, as it indicates the model's ability to identify and forecast network traffic accurately. The computation of the traffic discovery rate involves several steps. First, the model must be trained on historical network traffic data, which includes information on past network usage patterns, such as the volume and type of traffic at different times of the day. This data is used to train the model and build a predictive model that will be able to identify patterns and trends in the network traffic. Once the model is trained, it is tested on a separate dataset to evaluate its performance. **Fig 6** shows the comparison of traffic discovery rate

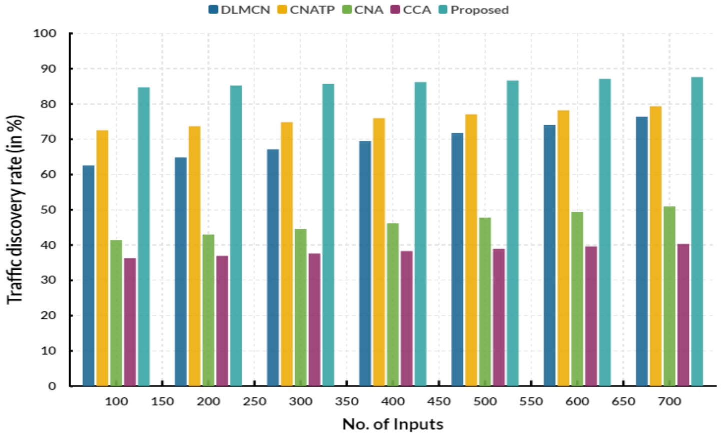

Fig 6. Comparison of Traffic Discovery Rate

This dataset contains network traffic data the model has not seen before, allowing for an unbiased assessment of its predictive capabilities. During the testing phase, the model will predict future network traffic based on the patterns and trends identified in the training data. The network traffic data is then compared to the model's predictions to determine its accuracy. The traffic discovery rate is calculated by dividing the number of correct predictions the model makes by the total number of predictions. It gives a percentage representing the model's accuracy in identifying and predicting network traffic patterns. A higher traffic discovery rate indicates a more accurate model, while a lower rate suggests that the model may need further refinement to improve its performance.

Computation of Congestion Discovery Rate

Congestion discovery rate (CDR) is a metric used to evaluate the effectiveness of predicting congestion in complex network traffic. It measures the proportion of observed congested periods correctly predicted by the congestion prediction model. The computation of CDR involves comparing the predicted congestion periods from the model with the actual congested periods observed in the network. The predicted and observed periods can be represented as binary vectors, where 1 indicates a congested period, and 0 indicates a non-congested period. **Fig 7** shows the comparison of congestion discovery rate.

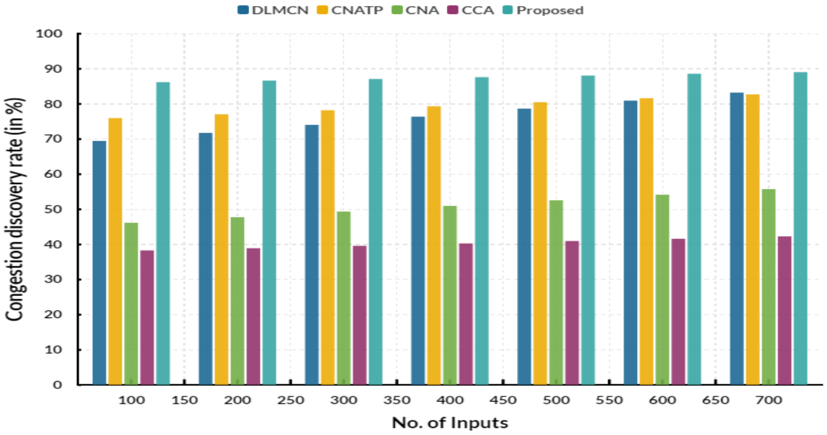

Fig 7. Congestion Discovery Rate

The first step in computing CDR is to determine the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) values. TP represents the correct predictions of congested periods, while FP represents the number of non-congested periods incorrectly predicted as congested. TN represents the correct predictions of non-congested periods, and FN represents the congested periods not predicted as congested. A high CDR indicates that the prediction model effectively identifies congested periods in the network. At the same time, a low CDR suggests room for improvement in predicting congestion. CDR can be used to compare the performance of different prediction models and identify the most accurate one for complex network traffic prediction.

V. CONCLUSION

Accurately predicting complex network traffic is challenging due to the dynamic and unpredictable nature of network traffic. However, advanced analytical tools such as deep learning and machine learning algorithms have shown promising results in forecasting network traffic patterns. These techniques can handle large and varied datasets and identify complex patterns and relationships in network traffic data. Moreover, the incorporation of real-time data and the continuous training of these models can improve the accuracy of predictions and adapt to changes in network traffic. Despite the potential benefits of these techniques, accurately predicting complex network traffic remains a challenging task that requires continuous research and development to keep up with the ever-evolving nature of networks. Furthermore, integrating these predictions into network management systems can significantly improve network performance and user experience. Therefore, it is crucial for network administrators to constantly explore and implement innovative techniques to predict and manage complex network traffic effectively.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

[1]. M. S. Sheikh and A. Regan, "A complex network analysis approach for estimation and detection of traffic incidents based on independent component analysis," Physica A: Statistical Mechanics and its Applications, vol. 586, p. 126504, Jan. 2022, doi: 10.1016/j.physa.2021.126504.

[2]. A. D'Alconzo, I. Drago, A. Morichetta, M. Mellia, and P. Casas, "A Survey on Big Data for Network Traffic Monitoring and Analysis," IEEE Transactions on Network and Service Management, vol. 16, no. 3, pp. 800–813, Sep. 2019, doi: 10.1109/tnsm.2019.2933358.

[3]. N. G. Álvarez, B. Adenso-Díaz, and L. Calzada-Infante, "Maritime Traffic as a Complex Network: a Systematic Review," Networks and Spatial Economics, vol. 21, no. 2, pp. 387–417, May 2021, doi: 10.1007/s11067-021-09528-7.

[4]. Z. Sui, Y. Wen, Y. Huang, C. Zhou, C. Xiao, and H. Chen, "Empirical analysis of complex network for marine traffic situation," Ocean Engineering, vol. 214, p. 107848, Oct. 2020, doi: 10.1016/j.oceaneng.2020.107848.

[5]. G. Wei, "Deep Learning Model under Complex Network and its Application in Traffic Detection and Analysis," 2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT, Oct. 2020, doi: 10.1109/iccasit50869.2020.9368560.

[6]. S. Ghosh, I. Saha Misra, and T. Chakraborty, "Optimal RSU deployment using complex network analysis for traffic prediction in VANET," Peer-to-Peer Networking and Applications, vol. 16, no. 2, pp. 1135–1154, Mar. 2023, doi: 10.1007/s12083-023-01453-5.

[7]. J. Zeng, Y. Xiong, F. Liu, J. Ye, and J. Tang, "Uncovering the spatiotemporal patterns of traffic congestion from large-scale trajectory data: A complex network approach," Physica A: Statistical Mechanics and its Applications, vol. 604, p. 127871, Oct. 2022, doi: 10.1016/j.physa.2022.127871.

[8]. Z. Tian, "Chaotic characteristic analysis of network traffic time series at different time scales," Chaos, Solitons &amp; Fractals, vol. 130, p. 109412, Jan. 2020, doi: 10.1016/j.chaos.2019.109412.

[9]. Z. Hu, F. Shao, and R. Sun, "A New Perspective on Traffic Flow Prediction: A Graph Spatial-Temporal Network with Complex Network Information," Electronics, vol. 11, no. 15, p. 2432, Aug. 2022, doi: 10.3390/electronics11152432.

[10]. Á. Fragua, A. Jiménez-Martín, and A. Mateos, "Complex network analysis techniques for the early detection of the outbreak of pandemics transmitted through air traffic," Scientific Reports, vol. 13, no. 1, Oct. 2023, doi: 10.1038/s41598-023-45482-9.

[11]. F. Catal, N. Tcholtchev, E. Höfig, and A. Hoffmann, "Visualization of Traffic Flows in a Simulated Network Environment to investigate abnormal Network Behavior in complex Network Infrastructures," Procedia Computer Science, vol. 151, pp. 279–286, 2019, doi: 10.1016/j.procs.2019.04.040.

[12]. F. Zhang, Y. Liu, L. Du, F. Goerlandt, Z. Sui, and Y. Wen, "A rule-based maritime traffic situation complex network approach for enhancing situation awareness of vessel traffic service operators," Ocean Engineering, vol. 284, p. 115203, Sep. 2023, doi: 10.1016/j.oceaneng.2023.115203.

[13]. X. Xin et al., "Multi-scale collision risk estimation for maritime traffic in complex port waters," Reliability Engineering &amp; System Safety, vol. 240, p. 109554, Dec. 2023, doi: 10.1016/j.ress.2023.109554.

[14]. X. Tao, Y. Peng, F. Zhao, S. Wang, and Z. Liu, "An Improved Parallel Network Traffic Anomaly Detection Method Based on Bagging and GRU," Lecture Notes in Computer Science, pp. 420–431, 2020, doi: 10.1007/978-3-030-59016-1_35.

[15]. M. Li, W. Yu, and J. Zhang, "Clustering Analysis of Multilayer Complex Network of Nanjing Metro Based on Traffic Line and Passenger Flow Big Data," Sustainability, vol. 15, no. 12, p. 9409, Jun. 2023, doi: 10.3390/su15129409.

[16]. K.-H. N. Bui, H. Yi, and J. Cho, "A Multi-Class Multi-Movement Vehicle Counting Framework for Traffic Analysis in Complex Areas Using CCTV Systems," Energies, vol. 13, no. 8, p. 2036, Apr. 2020, doi: 10.3390/en13082036.

[17]. Z. Sui, Y. Wen, Y. Huang, C. Zhou, L. Du, and M. A. Piera, "Node importance evaluation in marine traffic situation complex network for intelligent maritime supervision," Ocean Engineering, vol. 247, p. 110742, Mar. 2022, doi: 10.1016/j.oceaneng.2022.110742.

[18]. M. Abbasi, A. Shahraki, and A. Taherkordi, "Deep Learning for Network Traffic Monitoring and Analysis (NTMA): A Survey," Computer Communications, vol. 170, pp. 19–41, Mar. 2021, doi: 10.1016/j.comcom.2021.01.021.

[19]. W. Guan, H. Zhang, and V. C. M. Leung, "Analysis of Traffic Performance on Network Slicing Using Complex Network Theory," IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 15188–15199, Dec. 2020, doi: 10.1109/tvt.2020.3036934.

[20]. R. Ding et al., "Application of Complex Networks Theory in Urban Traffic Network Researches," Networks and Spatial Economics, vol. 19, no. 4, pp. 1281–1317, May 2019, doi: 10.1007/s11067-019-09466-5.

[21]. https://www.kaggle.com/datasets/crawford/computer-network-traffic