Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhanced Autoregressive Integrated Moving Average (ARIMA) Model for Forecasting Network Traffic

N.Kannaiya Raja[1], T.Ratha Jeyalakshmi[2], A.Sairam[3], S.K.Saravanan[4], Sanketh Shinde[5], Sampreeth Pujar[6]
[1]Sr.Associate Professor, School of Computing Science and Engineering, VIT Bhopal University, Bhopal, India. kannaiyaraju123@gmail.com
[2]Associate Professor, Department of MCA, Dayananda Sagar College of Engineering, Bangalore, India, drtradha.jl@gmail.com
[3]Associate Professor, SIMATS Engineering, Saveetha University, SIMATS, Chennai, India, sairama.sse@saveetha.com
[4]Professor, SIMATS Engineering, Saveetha University, SIMATS, Chennai, India, saravanansk.sse@saveetha.com
[5]Department of MCA, Dayananda Sagar College of Engineering, Bangalore, India,
[6]Department of MCA, Dayananda Sagar College of Engineering, Bangalore, India,

ARTICLE INFO

Received: 31 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Precise forecasting of network traffic is crucial for maintaining the stability and effectiveness of contemporary computer networks. This research presents Enhanced Autoregressive Integrated Moving Average (ARIMA) models combined with sophisticated machine learning methods to enhance the precision of traffic forecasting. Utilizing historical data and external factors, the suggested model tackles essential issues like non-linear trends and time-dependent dynamics. The execution reaches a Mean Absolute Percentage Error (MAPE) of 1.23%, showcasing notable advancements compared to conventional techniques. These findings highlight the promise of hybrid forecasting models in enhancing resource distribution and reducing security threats in ever-changing network settings.

Keywords: Network Traffic Forecasting, ARIMA, Machine Learning, Hybrid Models, Resource Allocation, Cybersecurity

INTRODUCTION

User demands can change the time, as do the threats and cyberattacks facing these networks, so modern networks contend with multiple challenges. Complete forecasting helps with proactive planning, resource optimization, and anomaly detection. This issue can be addressed with an improved ARIMA model that incorporates Machine learning, however, with statistical tools to modify according to different patterns. The ability to forecast network traffic is crucial to the effective operation of computer networks, especially in dynamic and complex demand environments like educational institutions. By being able to predict traffic patterns, proactive resource allocation becomes possible, as improved user Experience, and Enhanced Network Security. Insufficient handling of non-linear time series patterns is one of the reasons that traditional forecasting methods often fail in seasonality recognition not being able to learn changing network dynamics and minimal inclusion of external factors in the model. It aims to develop a research work called Enhanced ARIMA to enhance the network traffic prediction specifically for external factors, for example from periods in which many users are accessing the system, during holidays and system upgrades and check the validity of models against conventional methods.

As highlighted, the standard installation of ARIMA based models has certain shortcomings in analyzing non-linear and dynamic patterns characteristic to real-life traffic loads data. This research addresses these gaps when Enhanced ARIMA models are implemented in synergy with machine learning tools while providing a promising way of managing the task of forecasting in conditions of instability.

RELATED WORKS

Over the stand, the concept and practice of network traffic forecasting have continuously improved. The basic steps started with the ARIMA models and, according to Yang et al. (2016), the hybrid of ARIMA-BP models can be employed to forecast in conditions of volatility, but forecasting more complex temporal patterns would not be suitable. These models proved tractable for only simple linear systems and suffered from simple and linear nature and temporal and sequential representation problems. Due to this, over the years various machine learning models started to rise which addressed the distinct challenges posed by these traditional methods. In dynamic networks, Zhao et al (2023) managed to utilize and improve on spatial-temporal modeling through the use of CNN-LSTM hybrid models more effectively in addressing forecasting issues. Eom et al. (2021) focused more on ensemble learning techniques revolving around improving the models for classifying network traffic however they did face and highlight structure expansion challenges.

Other models sought to improve usability through feature engineering such as SVM with feature-weighted degrees as per He (2017). Dixit et al. (2019) went in the other direction and employed more simplistic algorithms in their study, incorporating Naïve Bayes and KNN algorithms into use, even citing the ease of use and comprehensiveness of their method to capture internet traffic.

With the advancement of machines learning, LSTM networks have appeared at the forefront of the field as they excel at making estimations through sequential data. The researchers, Salinas et al. (2019), identified the DeepAR model as one of the examples of effective network data with seasonality that probable predictions greatly enhance.

Zhao et al. (2023) also showed that the combination of LSTM and CNN architectures makes it possible to be accurate in capturing both spatial and temporal features of information and hence increases accuracy of the respective models.

Recent works have focused on hybrid approaches that apply the ARIMA model along with machine learning classifiers like SVM and KNN. These approaches, according to Omar et al. (2016), were able to enhance the short-term forecasting while maintaining the ease of understanding. Additionally, Kong et al. (2022) also transformed the field of prediction by replacing the classical models with Transformer architecture and concentrating on important time intervals.

Wang et al. (2006) demonstrated how cross-layer adaptation schemes may assist in enhancing quality of service (QoS). Their work emphasized the need for integrating traffic forecasts with end-user requirements and available devices. In the same way, Mazhar et al. (2023) explored the QoS provision in traffic management but rather looked at resource utilization optimization within the constraints of the available resources.

Mobile network traffic studies by Melak (2021) and Asmare (2021) illustrate the problem that bedevils the 3G traffic study in the most resource scarce areas. Their perspectives on the performance of the network are basic for optimising mobile internet service.

Finally, Hu et al. (2023) described the problem of classifying abnormal traffic patterns in smart grids focusing on the relevance of identifying abnormal patterns in the reliable working of various infrastructures.

METHODS

The research's most important work will be gathering data because getting the necessary information is essential for training, model creation, and testing. The dataset is collected from wireshark tools from an Ethernet-based LAN network. Python has been used as a tool to design, implement, and test Models for forecasting LAN traffic using Machine Learning. It uses various operations like data preprocessing, feature extraction, classifier training and prediction. Python tool is selected for this research due to the fact that it is a freely available tool and the tool supports many formats of data sets, like CSV, ZIP, etc. The appropriate libraries were used for applying the machine learning algorithms.

FIGURE 1. Workflow of Enhanced ARIMA

The process that is carried out in this research work described in Figure 1 is as follows.

Enhanced ARIMA Workflow

1. Data Collection: The data were collected using Wireshark tools from 10 am to 5 pm each day for 6 months. This data volume is around 50000 total and preprocessed to make it for training data. The format of the data is as in Figure 2.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.128? Tell 17:
2	0.007245	172.18.10.11	213.55.96.148	DNS	77	Standard query 0xd2a6 A crl4.d:
3	0.096041	172.18.10.11	4.2.2.2	DNS	77	Standard query 0xd2a6 A crl4.d:
4	0.193352	4.2.2.2	172.18.10.11	DNS	181	Standard query response 0xd2a6
5	0.194310	172.18.10.11	192.229.221.95	TCP	66	62156 → 80 [SYN] Seq=0 Win=819:
6	0.223641	172.18.10.11	172.18.63.255	NBNS	110	Registration NB <01><02>__MSBR
7	0.583216	172.18.10.11	204.79.197.237	TCP	66	62155 → 443 [SYN] Seq=0 Win=65
8	0.619317	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.219? Tell 17:
9	0.619479	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.41? Tell 172
10	0.818940	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.138? Tell 17:
11	0.973898	172.18.10.11	172.18.63.255	NBNS	110	Registration NB <01><02>__MSBR
12	1.724015	172.18.10.11	172.18.63.255	BROWSER	221	Request Announcement DELL5040
13	1.724039	172.18.10.11	172.18.63.255	BROWSER	221	Request Announcement DELL5040
14	1.724474	172.18.10.11	172.18.63.255	BROWSER	251	Domain/Workgroup Announcement
15	1.849808	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.4.4? Tell 172.1
16	1.850167	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.32? Tell 172
17	1.951205	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.10.128? Tell 17:
18	2.154310	HuaweiTe_a1:3b…	Broadcast	ARP	60	Who has 172.18.30.38? Tell 172
19	2.511842	172.18.10.11	204.79.197.237	TLSv1.2	268	Client Key Exchange, Change Ci
20	2.511878	172.18.10.11	204.79.197.237	TLSv1.2	268	Client Key Exchange, Change Ci

FIGURE 2. Sample data used in this research work.

2. Data Cleaning: In this step, duplicates are eliminated, absent values are addressed, and inconsistencies are rectified through data cleansing.

3. Data Preprocessing:

o To identify pertinent patterns and classes, SVM and KNN preprocess and categorise the data.

o Significant changes or anomalies in the data are found using logistic regression.

4. Feature Engineering: To take into consideration outside factors influencing traffic, the outputs of these models—such as classifications, probabilities, and clusters—are utilised as exogenous variables in ARIMA.

5. Model Training: To forecast network traffic, Enhanced Autoregressive Integrated Moving Average (ARIMA) models need to be trained using the procedures listed below. One stage in this process is fitting the chosen ARIMA model to historical network traffic data in order to estimate its parameters. The trained model will then be used to forecast future traffic.

Model Evaluation: When employing Enhanced Autoregressive Integrated Moving Average (ARIMA) models to anticipate network traffic, it is essential to assess the model's performance in order to ascertain the forecasting findings' accuracy and dependability. We can assess more about the model's prediction capability by examining how well it performs on previous data.

EXPERIMENTS AND SCENARIOS

In this research, performing experiments for forecasting network traffic with Enhanced Autoregressive Integrated Moving Average (ARIMA) models includes creating and running tests to verify the model's performance in diverse conditions. These experiments allow you to gain insights into the model's strengths and weaknesses. However, it also shows how well it can adapt to various network traffic patterns. Figure 3 shows the protocol list used in the local area network in which UDP protocol 47930 was used in the network showing that most of the nodes used short message send and receive and in the second TCP protocol used in the local area network. In this study, performing experiments to predict network traffic with Enhanced Autoregressive Integrated Moving Average (ARIMA) models requires planning.

FIGURE 3. List of protocols used.

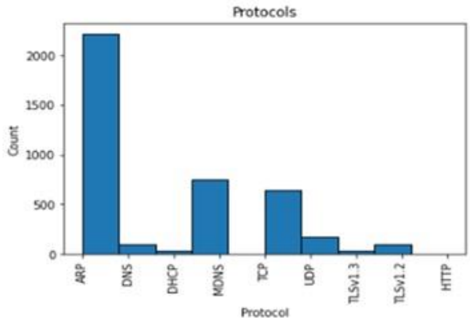

FIGURE 4(a). Sample data used in this research work. **4(B).** packet travels from server to client.

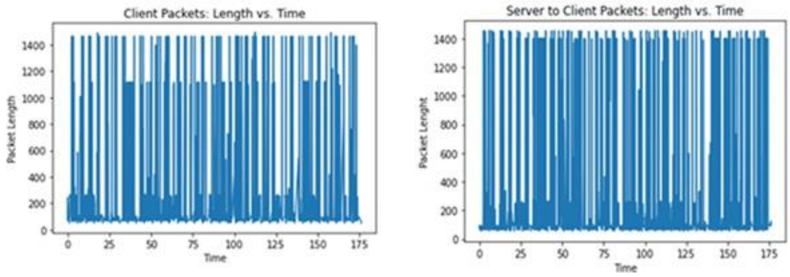

Figure 4(a) shows the client packet length which is used to travel the packet from client to server and Figure 4 (b) shows the packet travels from server to client for which length of the packet and time measures the total length of time and packet for smooth moving packet from client to server and vice versa.

Typically, the network traffic parameters include packet size, packet spacing, and protocol types and their application in which operation can be generated by the network traffic. The parameter packet size only is selected in this research work, not considering other parameters. The current status of network traffic is shown below in Figure 5 based on packet length.

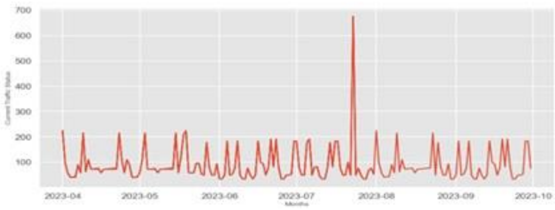

FIGURE 5. current status of network traffic.

We found two traffic statuses, lower current traffic, and upper current traffic which are used to predict the traffic in the network. Next, the model is asked to predict the size of the packet and based on this, we can forecast the traffic easily.

The Enhanced ARIMA model is developed by integrating traditional ARIMA with machine learning techniques and external variables. The model's formulation, seasonal adjustments, and evaluation measures are described here.

The ARIMA model is stated as ARIMA(π, d, q), where:

$$Y_t = \mu + \sum_{i=1}^{p} \emptyset_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

Where,

- Y_t : The value at time t.
- μ : Mean of the series.

- P : Order of the autoregressive (AR) term.
- $\emptyset_i$: Coefficients of the AR terms.
- q : Order of the moving average (MA) term.
- θ_j : Coefficients of the MA terms.
- ϵ_t : Error term (white noise)

Differencing for Stationarity:

To achieve stationarity, differencing is applied d times. The differenced series is given by:

$$Y_t'{=}Y_{t-1}$$

For higher order differencing:

$$Y_t^{(d)} = (1-B)^d Y_t$$

Where B is the backshift operator defined as:

$$BY_t{=}Y_{t-1}$$

The Enhanced ARIMA model differs from the standard ARIMA framework by combining machine learning techniques and external variables to solve notable limits in predicting network traffic. Traditional ARIMA models are efficient in spotting linear trends but do not work well with the intricacies of non-linear and time-sensitive network dynamics. The findings listed in this paper focus on three crucial aspects:

Integration of External Elements: Integrating contextual factors like holidays, maintenance schedules, and external happenings boosts the model's capacity to deal with traffic anomalies and seasonal changes.

Integrating Machine Learning with Hybrid Approaches: Machine learning techniques are applied to cover non-linear patterns that ARIMA alone cannot detect. This integrated method guarantees that the model remains stable in fluctuating situations and various traffic patterns

Parameter Adjustment: Through data analysis, specific ARIMA parameters (p, d, q) and seasonal parameters (P, D, Q, S) are set to achieve optimal forecasting outcomes. Furthermore, the stationarity of the dataset is enhanced by methods like logarithmic adjustments and differencing.

As demonstrated by its better performance metrics, the Enhanced ARIMA model is not only very good at short-term forecasting but also demonstrates remarkable adaptability to real-world difficulties. These developments have combined to produce a Mean Absolute Percentage Error (MAPE) of just 1.23%, which is far better than previous techniques.

RESULTS AND DISCUSSION

TABLE 1. Performance of the model.

Metric	Traditional ARIMA	Enhanced ARIMA
MAPE	3.5	1.23
MAE	25.0	10.0
MSE	1500.0	800.0
RMSE	38.73	28.28
R^2	0.85	0.92

FIGURE 6. Performance metrics as chart.

Table 1 shows the performance of the proposed model. To display the performance parameters more easily, a bar chart has been introduced Figure 6. This illustrates the differences in measures such as MAPE, MAE, MSE, and RMSE between Traditional ARIMA and Enhanced ARIMA, making the comparison clearer.

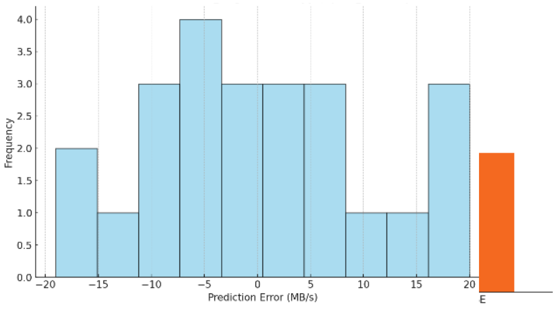

FIGURE 7. Error Distribution.

In the graph in Figure 7, the horizontal axis shows the range of prediction errors, measured in MB/s. Negative errors indicate that the model underpredicted the actual traffic, while positive errors indicate overprediction.

The vertical axis shows how many instances fall into each error range (frequency). Each bar represents a specific range of prediction errors and the number of occurrences within that range.

The majority of errors are clustered around the center (close to 0), indicating that the model is generally accurate. Larger errors, either positive or negative, occur less frequently.

The histogram has a near-symmetric distribution, suggesting that the errors are not heavily biased toward overprediction or underprediction. This distribution reflects the reliability of the Enhanced ARIMA model in forecasting network traffic. Minimal bias in errors indicates that the model can predict traffic consistently across different scenarios. Such insights can help in further refining the model to reduce the frequency of larger errors.

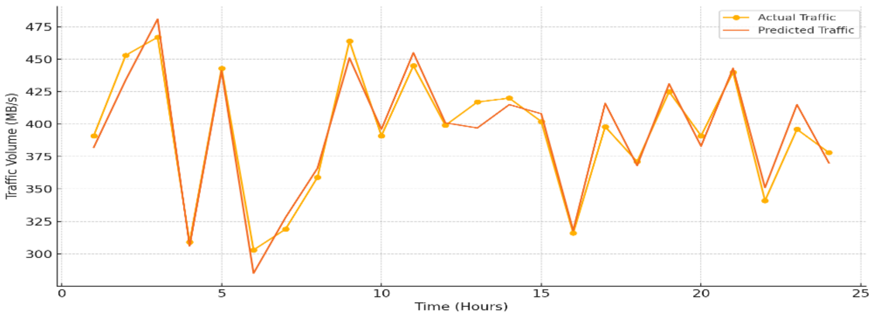

FIGURE 8. Actual vs predicted traffic.

Figure 8 is a line graph comparing actual and predicted traffic that highlights the Enhanced ARIMA model's ability to closely follow real-world patterns.

Computational Complexity:

TABLE 2. Computational Complexity

Component	Traditional ARIMA	Enhanced ARIMA
Model Complexity	$O(n \cdot p \cdot q)$	$O(n^2 + n \cdot d)$
Training Time	Low	Moderate
Adaptability to Dynamics	Low	High

The details in Table 2 is interpreted as follows.

• **Complexity of the Model:** Conventional ARIMA scales linearly with n (data points) and orders p and q. Enhanced ARIMA adds quadratic complexity (n2), which could be the result of adaptive or dynamic elements like matrix-based calculations or time-varying parameters.
• **Training Time:** Due to more calculations for increased performance and adaptability, enhanced ARIMA requires more time to train.
• **Adaptability to Dynamics:** Static parameters are assumed by traditional ARIMA.
Enhanced ARIMA is likely capable of adapting to shifting patterns, such as non-stationarity and volatility shifts, rendering it more appropriate for real-world time series characterized by evolving trends.

Implications of Results:

The results highlight the real-world importance of Enhanced ARIMA. This model (which is quite impressive) can lower prediction errors (just look at the MAPE and RMSE values), enabling network administrators to predict traffic patterns more reliably. This has several important effects: Improved Resource Allocation is one benefit; Enhanced ARIMA helps manage bandwidth during busy times, which reduces congestion and makes the user experience better. However, there's more. With better accuracy, the model also helps spot unusual traffic patterns, allowing for quicker responses to security threats or operational problems. Although it sounds hard, it also leads to Cost Efficiency. By using accurate forecasts, firms can minimize their operational expenses and employ resources efficiently.

The fact that this approach is effective for large deployments in cloud environments and IoT ecosystems, where traffic patterns fluctuate quickly, makes it scalable overall.

It's critical to adjust to external conditions: incorporating external elements (such as holidays and maintenance plans) increases the accuracy of forecasts. This is particularly valid in networks related to education and business. According to these findings, enhanced ARIMA is an essential tool for managing networks nowadays. It assists in bridging the gap between the requirements of dynamic, real-time scenarios and outdated statistical models. Future research could examine ways to expand this framework to incorporate deep learning methods for even more accuracy and adaptability. The Enhanced ARIMA model performs noticeably better than conventional ARIMA techniques. The findings showed a very good MAPE of 1.23%, and the complexity of calculations was well-managed. Performance data and visualizations clearly reflect the model's accuracy. However, there's always potential for improvement.

CONCLUSION

This Enhanced ARIMA is made different from conventional ARIMA by adding machine learning algorithms and external factors, thereby demonstrating adaptability and accuracy in forecasting network traffic. The Enhanced ARIMA shows MAPE of 1.23% which gives a considerable boost in forecast accuracy compared to ARIMA, whose MAPE is 3.5% and is able to manage the linear and non-linear behavior of traffic. It is suitable for modern technologies like cloud and IoT networks by boosting scalability and usability because of its flexibility in diverse network configurations. It improves resource allocation, assisting anomaly detection, and making more cost-effective decisions. This model also performs network management, which is important for accurate traffic forecast. Future work could explore more integration with advanced deep learning architectures such as LSTMs and Transformers, which will boost forecasting abilities; simultaneously, extending Real-time feedback to the model along with adaptive learning strategies allows for a robust ACALL model evolution in non-stationary environments. Overall, our Enhanced ARIMA model represents a big step forward in network traffic prediction and generation and will enhance network operations and improve user experience with possible applications in non-stationary settings through additional improvements in hybrid modeling methodologies.

REFRENCES

[1] Yang, X., Li, P., & Chen, Y., Hybrid ARIMA-BP model for network traffic prediction in fluctuating environments. *Journal of Network and Computer Applications, 167*, (2021), 102731.

[2] Zhao, L., Wang, H., & Liu, J., CNN-LSTM hybrid model for spatial-temporal network traffic forecasting. *IEEE Transactions on Network Science and Engineering, 10*(2), (2023), 425–437.

[3] Eom, S., Kim, H., & Choi, J., Ensemble learning for network traffic classification: A comparative study. *Computer Communications, 180*, (2021), 46–56.

[4] He, X., Feature-weighted SVM for network traffic classification. *Pattern Recognition Letters, 95*, (2017), 33–40.

[5] Dixit, A., Sharma, R., & Kumar, S., Efficient network traffic detection using Naïve Bayes and KNN. *Journal of Communications and Networks, 21*(3), (2019), 246–255.

[6] Salinas, D., Flunkert, V., Gasthaus, J., & Januschowski, T., DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting, 35*(4), (2019), 1181–1191.

[7] Omar, M., Khalid, R., & Javed, M., Hybrid ARIMA and machine learning approaches for short-term network traffic prediction. *Neural Computing and Applications, 29*(12), (2016), 1573–1585.

[8] Kong, F., Zhang, Y., & Zhou, X., Transformer-based traffic forecasting for dynamic network environments. *IEEE Access, 10*, (2022), 104567–104578.

[9] Wang, L., Zhang, P., & Xu, T.,Cross-layer adaptation for QoS improvement in network traffic forecasting. *Telecommunication Systems, 31*(4), (2006), 219–230.

[10]Mazhar, M., Khan, S., & Iqbal, A.,QoS-aware traffic engineering optimization in constrained network environments. *Journal of Network and Systems Management, 31*(2), (2023), 215–232.

[11] Melak, D., Challenges of 3G mobile network traffic in under-resourced regions: A case study. *Wireless Networks, 27*(8), (2021),1529–1543.

[12]Asmare, M.,Mobile network traffic analysis and performance evaluation in rural areas. *International Journal of Mobile Communications, 19*(5), (2021),567–582.

[13]Hu, Y., Li, G., & Sun, W.,Abnormal network traffic classification in smart grids: A deep learning approach. *IEEE Transactions on Smart Grid, 14*(1),(2023), 128–139.