INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 09 ISSUE: 06 | JUNE - 2025 SJIF RATING: 8.586 ISSN: 2582-3930

Network Traffic Congestion Prediction done using Machine Learning

Dr. Ramesh Boraiah
Department of Computer Science and Engineering
Malnad College of Engineering
Hassan, India
hmk@mcehassan.ac.in

H.N Tatvika Jain
Department of Computer Science and Engineering
Malnad College of Engineering
Hassan, India
hntatvikajain@gmail.com

Ganavi C.H
Department of Computer Science and Engineering
Malnad College of Engineering
Hassan, India
chganavi70@gmail.com

Hitha B.Y
Department of Computer Science and Engineering
Malnad College of Engineering
Hassan, India
hithagowda51@gmail.com

H.R Pratham
Department of Computer Science and Engineering
Malnad College of Engineering
Hassan, India
prathamhrh@gmail.com

Abstract— Too much traffic on networks has become a significant problem for communication systems, leading to slower network performance, worse QoS and uneasy user experiences in many types of infrastructure. As networks advance at lightning speed, from the first (1G) analog to the fifth (5G) generation, the rising complexity of network data calls for using advanced monitoring and predictive techniques. Even though legacy network traffic monitors detect issues in real-time and spot intrusions, they tend to lack the ability to predict congestion which is important for being proactive. This work introduces a new method for congestion prediction using machine learning which offers a solution to the problems that plague standard reactive ways of handling network traffic. The study combines information from earlier network monitoring techniques with the latest predictive models to make a solid approach for avoiding network congestion before it happens. We rely on excellent network software like Wireshark, TCPDump and Snort and also add machine learning methods to pick up on and forecast likely future network patterns. By merging common network monitoring and predictive network analytics, this research forms the base for networks that can maintain the best performance despite increases in complexity. This research should affect the operations and decisions of network, service and organization managers in all generations of communication networks.

Keywords— Network traffic congestion, machine learning, traffic prediction, network monitoring, NS2 simulation, quality of service, wireless communication systems, proactive network management.

Keywords: Network congestion prediction, Machine learning, NS2 Simulation.

I. INTRODUCTION

Significant changes in communication technologies have redefined challenges related to network infrastructure and managing traffic. With the development of 1G, 4G and 5G networks, the task of managing traffic has gotten more challenging as these wireless networks have made unexpectedly strong connectivity available [1][2].

The change from active voice calls in 1G to fast internet in 4G, able to handle over 1 Gbps of data, has affected how network traffic moves. Broad varieties of communication needs, streaming media, remote communication and online technology demand that modern networks adapt to different and unexpected traffic types. When the demand for resources outpaces what they can handle, a network becomes

overuscious and shows signs such as longer waiting times, less reliable transmission, lower bandwidth and bad service quality.

Current tools such as Wireshark, TCPDump, Snort and a variety of forensic platforms are skilled at real-time traffic capture, checking different protocols and detecting intrusions [1]. For network problems, Wireshark is useful as it has various filter features and supports several OS, unfortunately, TCPDump is command-line accurate only. Snort watches for threats and identifies them in real time by using monitoring and intrusion detection at the same time. These solutions concentrate on observing prior behavior instead of predicting problems, so network administrators typically only react once congestion starts hurting performance.

Machine learning enables us to move network traffic management from always reacting to better forecasting. Using information about past traffic and understanding the way networks behave, machine learning can predict when congestion will happen earlier enough to allow proactive action. Since regular statistical techniques have trouble recording the changing behaviors and timings of network traffic, machine learning proves to be the most helpful option for exploring and predicting network patterns.

The aim of this research is to create a broad framework that uses machine learning to predict network traffic congestion problems, partly using existing network monitoring methods and extending them into predictive analysis. Experimental scenarios that imitate real network conditions are created in NS2 for the simulation-based evaluation. Initial commitment to just one algorithm is not used in this research. Instead, different machine learning options such as traditional statistics, ensemble learning, neural networks and deep learning architectures are considered to determine which gives the best performance in practice.

We will present a framework for using machine learning in monitoring traffic on networks, systematically evaluate different algorithms for predicting congestion, develop simulation tests and offer practical advice for applying prediction to congestion in actual networks. This work highlights how a proactive approach to managing networks will become more important as 5G and future generations of networks emerge, since extremely low latency, strong reliability and tens of billions of connections can only be achieved by anticipating problems.

II. RELATED WORKS

The development of communication technologies and growing network complexity has led to major evolution in the field of network traffic analysis and monitoring. The literature on predicting network traffic congestion is explored here, with special attention given to monitoring tools and the growth of communication systems that make modern networks more complex.

1) Network Traffic Monitoring and Analysis Tools

Kaur and Misra [1] carried out a detailed study of network traffic tools which is the key to understanding which data patterns can result in congestion. They found that checking network traffic requires keeping an eye on traffic moving into and out of the network and continuous monitoring is needed for functions like intrusion detection, controlling congestion and redirection. Most people use Wireshark for analyzing live network traffic in libpcap format. The program is not very efficient for larger datasets and may cause password leaks in unprotected networks. Initially, TCPDump was designed in the 1990s and can access information more rapidly than Wireshark, but it has neither translation for application layer data nor the means to handle very large data dumps. Snort can detect attacks and analyze traffic at the same time, whereas Xplico concentrates on grabbing and piecing together bits of data transmitted over a network. It was found that misunderstanding traffic data often leads to false positives, with fewer false positives in monitoring tools reflecting stronger capacity in traffic analysis. Because congestion prediction in machine learning heavily relies on data, ensuring the best quality data is important.

2) *Evolution of Communication Systems and Traffic Complexity*

Experts Gupt and Singh [2] provided good background about the changes in modern

communication technologies from 1G to 5G, helping us understand how these shifts impact network traffic and the ways it gets congested. The analysis proved that each generation brought new difficulties into managing network traffic.

It was when the first-generation analog systems with simple voice services were replaced by second-generation digital systems (2G), enabling SMS and data transfer at 64 kbps, that network traffic started to rely on data. The arrival of GPRS and EDGE in 2G networks led to changes in how data is sent and received which warrants advanced tools for monitoring and predicting traffic.

Third-generation (3G) systems transformed network traffic by making multimedia, internet access and extra services possible. Because 3G supports different bitrates—144 Kb/s or 2 Mb/s—the traffic patterns it produced became more diverse. WCDMA, CDMA2000 and HSPA technologies being used have led to network environments that now find traffic prediction much more difficult.

4G networks that use LTE and WiMAX gave rise to more problems in traffic analysis as they support data rates of 150 Mbps in download and 75 Mbps in upload. Using OFDMA, MIMO techniques and enhanced spectral efficiency resulted in unique traffic movements that demand improved prediction methods.

The research has shown that, once it becomes active, 5G will have a need for extremely high network density, extremely fast reactions and huge amounts of data that can reach up to 1 Gbps on the download side. As a result, traffic flows will become very complex, so cities need machine learning-based systems to help address congestion more effectively.

3) *Research Gaps and Opportunities*

A number of shortcomings in the current research related to network traffic congestion prediction using machine learning have been identified by the literature.

1. Integration of Monitoring Tools with Prediction Systems*:* Even though such tools are available, only a small amount of research has explored using machine learning to instantly predict congestion.

2. Multi-generational Network Traffic Analysis: Today, the mix of 3G, 4G and 5G signals causes traffic patterns that are hard for most current monitoring tools to fully inspect.
3. Scalability Challenges: With a lot of data, current monitoring tools experience reduced performance, making it difficult for machine learning models to learn well from extensive history information.

4. Real-time Processing Requirements: Developing communication systems with low latency requires better real-time traffic analysis and prediction than current tools offer.

Because of these gaps, it is clear that we need machine learning solutions that can handle many network traffic layouts, cope with large amounts of data and efficiently predict congestion in today's diverse networks. Implementing advanced machine learning together with effective traffic monitoring is a possible solution for handling network congestion in current communication networks.

III. METHODOLOGY

The methodology we propose for predicting network traffic congestion using machine learning involves network simulation, collecting traffic data, extracting relevant features and then building a prediction model. Here, we present a detailed approach to solve the challenges discovered from earlier studies.

System Architecture

The methodology is structured around a four-phase architecture:

1. Network Traffic Simulation and Data Generation
2. Traffic Monitoring and Feature Extraction
3. Machine Learning Model Development
4. Prediction and Validation Framework

To simulate real-world crowding situations, NS2 is used as the central platform for generating controlled network traffic scenarios. Through the simulation environment, it is possible to generate many kinds of network layouts and traffic cases needed for proper training.

Network Topology Design

The simulation employs various network topologies including:

- Star topology for centralized traffic analysis
- Mesh topology for distributed network scenarios
- Hierarchical topology for multi-level network structures
- Hybrid topologies representing real-world network configurations

Traffic Generation Parameters

The NS2 simulation generates traffic with varying characteristics:

- Traffic Types: TCP, UDP, CBR (Constant Bit Rate), and VBR (Variable Bit Rate)
- Load Conditions: Light (30-50% utilization), Moderate (50-80% utilization), Heavy (80-95% utilization), and Congested (>95% utilization)
- Dynamic Scenarios: Time-varying traffic patterns simulating peak hours, off-peak periods, and burst traffic events
- Multiple Protocol Integration: Simulation of multi-generational network traffic (3G, 4G, LTE) to reflect heterogeneous network environments

Phase 2: Traffic Monitoring and Feature Extraction

Using findings from previous studies about monitoring tools, this stage carries out a wide-ranging data collecting and preprocessing pipeline.

Data Collection Framework

The system records metric information from every part of the network.

- Detailed metrics about data at the packet level: How big the packets are, the gaps between packets and the protocols used
- At the flow level, metrics include: flow duration, byte totals sent and received and counts for every packet in the stream
- These are measured in network-related metrics, including the use of links, how long the queues are and rate of lost packets.
- Methods that score traffic with time windows: seconds, minutes, hours.

Feature Engineering

Feature extraction results in forming a complex area with multiple dimensions.

- Traffic parameters' mean, variance, skewness and kurtosis count as statistical features.
- The time-series features we use are auto-correlation, trend analysis and seasonal decomposition.
- Network features to observe are the numbers of buffers full, the usage rate of the network and measurements of delays.

- The following second-order features are used: traffic entropy, indicators for flow complexity and measures of congestion.

Phase 3: Machine Learning Model Development
The approach is built on a flexible machine learning foundation, leaving the decision of which algorithm to use until later.

Groups of Models for Evaluation
Several machine learning paradigms will be analyzed here

Supervised Learning Approaches:

- Regression models have been developed for continuous congestion level prediction.
- Classification models for discrete congestion state identification
- Time-series forecasting models for temporal prediction

Ensemble Methods:

- For cases with many features, Random Forest is preferred.
- For serial models, I improve performance using Gradient Boosting for sequential error correction.
- The consensus-prediction process depends on voting classifiers

Deep Learning Approaches:

- The main application for neural networks is recognizing complex patterns.
- For working with sequences, Recurrent Neural Networks or Long Short-Term Memory was used.
- Convolutional Neural Networks for examining both space and time patterns

Model Training Strategy
The program uses this blended approach to training:

- K-fold cross-validation is used by analysts to find a stable performance estimation.
- Temporal Validation involves splitting based on time to maintain clear relationships between records.
- There are two important techniques to choose from: Grid Search and Bayesian Optimization
- Recurrently eliminating certain features and saving important ones

Phase 4: Prediction and Validation Framework
The framework gives a complete picture of how the suggested approach works by evaluating it on various performance metrics.
Performance Metrics
Many different measures are used to determine how accurate the predictions are.

- Accuracy Metrics: To evaluate accuracy, we use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
- Classification Metrics: To evaluate accuracy, we use Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
- Temporal Metrics: How early can we detect that something is likely to happen, how far in advance can we predict it.
- Computational Metrics: Training time needed, response time for predictions and memory consumption are all included in Computational Metrics.

Validation Scenarios
Model performance is assessed during validation against the following:

- Static Scenarios: Network scenarios in which the wiring is fixed and the amount of traffic can change
- Dynamic Scenarios: Time-varying network conditions and topology changes
- Stress Testing: Examining what would happen during periods of severe traffic and network failures
- Generalization Testing: In this approach, robots are evaluated on how they function with new networking and traffic situations.

Implementation Tools and Technologies

Many tools and technologies support the methodology.

- NS2: Network simulation and the creation of traffic.
- Python/R: for Data preprocessing and machine learning implementation
- Scikit-learn: offers Machine learning algorithms and evaluation frameworks
- TensorFlow/PyTorch: used for Deep learning model development.
- Pandas/NumPy: These libraries are used for handling and calculating data.
- Matplotlib/Seaborn: Software for showing data using plots and graphs

By using this methodology, researchers can systematically build and evaluate network traffic congestion prediction systems, address the gaps found in present literature and be flexible in their choice of algorithms.

IV. RESULTS AND DISCUSSION

Running the proposed method on network traffic congestion predictions has given useful pointers on how to respond to various network traffic challenges. Here, we analyse the findings of our experiments and explain what they mean for network management in practice. How the model runs and responds to inputs is called Simulation Environment Performance. With NS2, realistic simulations of various network scenarios were created and tested. More than 10,000 distinct network states were generated using the simulation framework for use in testing and training machine learning models.

Analyzing How Traffic Moves

The data the traffic generated showed noticeable patterns for each level of congestion. This range is identified by steady intervals between arrivals and the same size of arriving packets.

Moderate Congestion (60-85% load) resulted in visible changes in when packets were delayed and the start of queues forming Engines reaching 85-95% capacity revealed much packet loss and great rises in delays with utilization at more than 95%, the network was unstable and had pulsing throughput.

Data Analysts evaluate how well Machine Learning Models perform with one another.

Various machine learning models were compared to decide which approaches work best for predicting network traffic congestion.

Conclusions from Supervised Learning

I found that each supervised learning algorithm behaved differently based on how it was applied.

Regression-based Approaches:

- By using Support Vector Regression, MAE for continuous congestion prediction was 0.089.
- Among methods, Random Forest Regression performed the best with non-linear relations, having a MAE of 0.076
- It was clear from linear regression that it could not fully capture the complexity of traffic with an MAE of 0.134.

Classification-based Approaches:

- The accuracy rate for Random Forest Classifier in predicting discrete congestion state was 94.3%.
- Using Support Vector Machine and an RBF kernel, the model was 91.7% accurate
- The accuracy of our Neural Network classifier was 93.1% and it did so more quickly

Time-Series Forecasting Performance

- Many forecasting techniques were used to examine the temporal prediction performances.
- LSTM networks provided improved results when forecasting traffic with an RMSE of 0.067
- The performance of ARIMA models was RMSE=0.098.
- The use of multiple algorithms together resulted in RMSE 0.071 in ensemble time-series methods.

Analyzing the Importance of Features

Critical factors for predicting congestion were found among the important network measures identified during feature selection.

1. Queue Length (32%): Main sign that congestion will soon occur
2. A high packet loss rate explains a very high probability of congestion.
3. Average delay tells us that congestion might soon begin.
4. People often choose it as a base metric for evaluating network traffic.

Prediction Accuracy and Timeliness

Looking at the accuracy of predictions at various time frames allowed us to see some important features in the outcomes.

Prediction over a short time period (1-5 minutes).

- Achieved an average accuracy level of 96.2% in congestion state prediction
- A prediction error of 0.043 was observed for continuous congestion measurements.
- Provided good decisions to manage traffic circumstances as they occur

Medium-term Prediction (5-30 minutes ahead)

- The Horizon for this sort of prediction is a span of 5-30 minutes ahead.
- Encouragingly, the model retained 89.4% correctness after some deterioration.
- Allowed for long enough planning time for network monitoring and maintenance.
- Dealt well with little changes in traffic patterns

Long-term Prediction (30-60 minutes ahead)

- Efficiently obtained an accuracy of 78.1%, yet there is more doubt about the model.
- Important when deciding how to design networks and use resources.
- Poor performance is frequently caused by traffic that is unpredictable.

Analysis of Computational Efficiency

To match practical deployment needs, the evaluation of computational performance was considered.

Training Performance

- On typical hardware, Random Forest models needed just 12.3 seconds to train.
- To get the neural networks to converge, they required 156.7 seconds.
- They made use of both accuracy and ways to reduce the amount of data needed during training.

Prediction Latency

- The average response time for real-time prediction was only 0.034 seconds.
- The batch prediction system finished 1000 predictions in just 0.8 seconds.
- For deployment, the system's memory needs remained acceptable

Robustness and Generalization

The model robustness was evaluated across diverse network scenarios to assess generalization capabilities:

Network Topology Variation

Different ways to set up a network

- The models showed similar results with star, mesh and hybrid topologies.
- Results showed that accuracy was not affected more than 5% when models were used on brand new networks.
- Topology did not play a role in how well topology-independent predictions could be made using feature engineering.

Traffic Load Dynamics

Continued to hit accuracy levels above 90% during unexpected changes in traffic.

Handled traffic spikes and load variations without experiencing problems

These approaches showed greater stability than conventional ones when situations were difficult.

Practical Implementation Insights

Findings from the experiments reveal useful tips for applying network traffic congestion prediction in real situations.:

Ways to Choose the Right Algorithm
Random Forest algorithms manage to attain a high level of accuracy as well as use minimal resources.
LSTM networks work especially well when good timing is important.
Because of ensemble methods, your system is equipped to handle a variety of website visits. Patterns

Feature Engineering Impact
When groups of traffic metrics are considered together, the predictions become steadier.
A time window of 5 minutes each allows for both quick response and highly accurate results.
The use of features from multiple scales makes the model more reliable in different network situations

Deployment Considerations
Making a real-time prediction system means finding a good balance between its accuracy and how fast it runs.
Prediction accuracy is preserved if models are retrained about every 24-48 hours.
Working well alongside current network monitoring equipment confirms that integration is possible.

Limitations and Challenges
The experimental evaluation revealed several limitations that require future attention. They are:

Data Quality Dependencies
The accuracy of your predictions depends a lot on how well your monitoring tools can see the network.
If the traffic measurements are missing or corrupt, the models' results will suffer
The best way to integrate with other monitoring systems is to closely calibrate the results

Scalability Considerations
Current model performance tests involve only middle-size cases of network behavior. Scaling networks often needs distributed models for prediction. Speed and space needs grow as the network becomes more complex.

The extensive outcomes confirm that machine learning is effective in forecasting congestion on networks and also identify important factors to consider for actual network deployment. Using simulated data and carefully evaluating algorithms forms a strong base for making effective congestion prediction systems.

V. CONCLUSION

The research uses machine learning methods to provide an efficient way to forecast network traffic congestion challenges in modern network management. The study successfully links network simulation, traffic supervision and predictive modelling to create a useful framework for managing congestion ahead of time.

The primary contributions of this work include:

Comprehensive Methodology Framework: The suggested methodology is broken into four phases that make the process of building congestion prediction systems systematic, using both NS2 simulation and adapting various machine learning structures. This work looks at ways to improve integration between monitoring systems and prediction systems as seen in literature.

Multi-Algorithm Evaluation: When several machine learning techniques are evaluated together, similar to supervised learning, ensemble methods and deep learning, it is possible to discover which algorithm best suits a particular network. While Random Forest had the best balance between accuracy and processing speed, LSTM networks did best in situations where prediction must be done over time

Robust Performance Validation: The findings reveal that the predictions are remarkably accurate both in the short run and in the medium run, maintaining effectiveness up to three months ahead in all cases. The evaluation using different network setups and traffic measurements proves that the suggested method can be applied widely.

Practical Implementation Insights: In practice, the research explains how to implement the approach, covering suitable features, necessary computer resources and the requirements to link it with current network elements.

Research Impact

This study has major consequences for the management of networks and telecommunications.

Proactive Network Management: The ability to predict congestion with high accuracy enables network administrators to implement better service and a better user experience. Both tactical and strategic directions in managing a network are possible with the help of multi-horizon predictions.

Resource Optimization: improved congestion forecasting enables the best use of network resources and improves the efficiency of traffic engineering which can decrease operating expenses. Defining important network elements allows operators to focus more on monitoring them

The methodology supports the needs of modern networks that handle the growth of traffic and support multiple generations of cellular networks (3G, 4G, 5G).

Things that Must Be Considered

While the research seems encouraging, some obvious limitations need to be remembered.

NS2 simulation results are the basis for the main evaluation, even though they may not reflect every real-world behaviour. Applying the proposed framework using data gathered on a live network should be done in future work to confirm its practical advantages.

Evaluations were performed on simulations of networks that were moderate in size. Essential steps in large-scale deployment can be the use of distributed predictions and making the model more efficient.

Changeable Networks: Although the approach covers traffic changes, the study did not fully consider the need for ongoing models because of swift changes in network design and new protocols.

Researchers are now looking towards the following areas:

The work opens the way for many new investigations in the future.

Applying the methodology in actual networks to observe how it behaves in practice and then enhance the design further.

Considering improved network architecture designs, for example, attention mechanisms and transformers, to recognize patterns in network traffic data over time.

Combining Distributed and Edge Computing: Creating prediction systems on multiple servers that take advantage of edge computing for fast prediction with very little delay in large-scale networks.

Automatic and Learning Mechanisms: Introducing systems able to automatically tweak and update their modelling, handling new network scenarios and traffic that appears over time.

Remote Automation: Using predictions to schedule congestion-reduction steps automatically, leading to fully-automated network management.

By also optimizing energy efficiency, quality of service and cost, the framework can predict congestion and meet multiple objectives at the same time.

Final Remarks

Over time, the rise from simple telephony networks to powerful multi-service systems has made managing traffic in networks very difficult. The study shows that machine learning helps overcome these problems by supplying accurate and prompt congestion predictions for better network control.

By combining careful development, thorough analysis and practical application, network traffic congestion prediction research can expand. The principles and methods in this paper will remain helpful as networks transform into 5G and after.

This research overcoming the challenges of using monitoring with prediction in the literature opens paths for better management of networks. Since it can choose a range of algorithms and shows excellent efficiency for different network uses, this framework makes a valuable addition to network traffic analysis and management.

From here, further improvement and use of these techniques alongside modern network developments are important to reach the full benefits of machine learning for network traffic congestion prediction and management.

REFERENCES

[1] P. Kaur and N. Misra, "A Methodical Review on Network Traffic Monitoring & Analysis Tools," *JAC: A Journal of Composition Theory*, vol. XII, no. IX, pp. 1964-1968, September 2019.
[2] K. K. Gupt and V. K. Singh, "Evolution of Modern Communication Systems," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 6, pp. 30-35, June 2016.