Traffic Prediction for Research and Education Networks using an Ensemble GRU-LSTM with Varying Lead Times

Mohammad Arafath Uddin Shariff, Venkat Sai Suman Lamba Karanam, and Byrav Ramamurthy
School of Computing
University of Nebraska-Lincoln
Lincoln, NE, USA
mshariff2@unl.edu, saisuman1@acm.org, byrav@unl.edu

2025 IEEE International Conference on Communications Workshops (ICC Workshops) | 979-8-3315-9624-8/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICCWorkshops67674.2025.11162466

***Abstract*—Research and Education Networks (RENs) are critical infrastructures that facilitate high-bandwidth data transfers for scientific collaborations and educational activities. While numerous traffic prediction techniques exist for public Wide Area Networks (WANs), they often fail to capture the unique characteristics of REN traffic. This paper presents a novel approach to traffic forecasting tailored for RENs, using real-world traffic traces from Internet2, a major REN in the US. We leveraged these traces to develop a high-performance forecasting model and provide key insights into REN traffic behavior. Our main contributions are as follows: (i) We performed an in-depth analysis of core router traffic from the Internet2 backbone, revealing distinct cyclical patterns that are highly amenable to prediction. Specifically, we examined traffic from two core routers, identifying unique temporal dynamics. (ii) We designed an innovative hybrid Recurrent Neural Network (RNN) model, combining Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) layers (GRU-LSTM), to effectively capture both long-term dependencies and short-term fluctuations inherent in REN traffic. (iii) We demonstrate that the prediction *lead time* is a critical *hyperparameter* for REN traffic forecasting, significantly impacting model accuracy. This contrasts with typical WAN traffic patterns and highlights the unique predictability of REN traffic due to consistent, experiment-driven flows. (iv) We propose to enhance our traffic prediction model by integrating traffic anomaly detection, thereby improving its robustness against unexpected traffic surges.**

***Index Terms*—Internet2, Network Traffic Analysis, Anomaly Detection, Machine Learning, LSTM, GRU, Prophet, Isolation Forest**

I. Introduction

Research and Education Networks (RENs) are vital infrastructures for data-intensive scientific and academic collaborations, facilitating high-speed data transfers essential for breakthroughs in fields like high-energy physics and genomics. Accurate REN traffic forecasting is crucial for optimizing network resources and ensuring efficient operations. While extensive traffic forecasting research exists for general Wide Area Networks (WANs), Research and Education Networks (RENs) with unique traffic characteristics are often overlooked. REN traffic, driven by scheduled academic activities and large scientific data transfers, differs significantly from commercial internet traffic.

Existing forecasting techniques, including Deep Learning (DL) models, primarily target general-purpose WANs or IoT networks [1]–[3]. Although some application-level forecasting exists [4], [5], network-level REN traffic forecasting remains largely unaddressed. In a recent study [6], a deep matrix factorization approach for large-scale modeling of Internet backbones was presented, emphasizing the challenges of capturing diverse traffic patterns at scale. However, we bridge this gap by presenting novel ML-based forecasting methods for REN traffic. The scarcity of network-layer REN datasets poses a significant challenge, especially when dealing with large-scale or big data scenarios [7]. Through collaboration with Internet2 [8], a major US REN, we collected network-layer traffic data from core routers, providing valuable insights for analysis and model design.

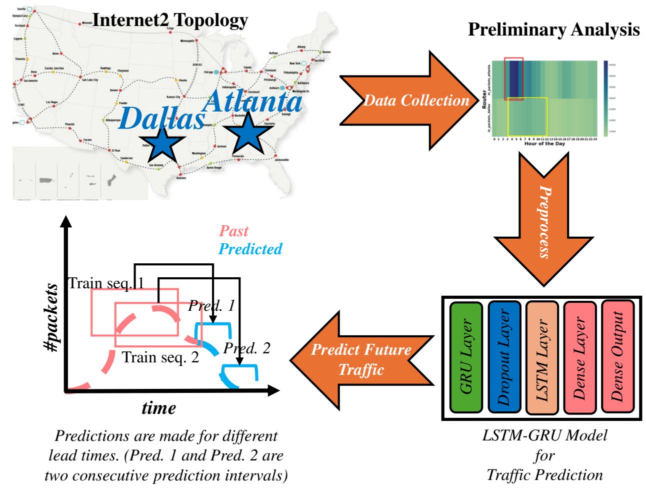

Fig. 1. High-level architecture of our traffic prediction.

This paper focuses on accurately forecasting network traffic volume in RENs. Given Internet2 core router traffic data, we aim to develop a model to predict hourly incoming packet counts (*#Packets*). Accurate traffic prediction is essential for REN resource management and performance. To ensure research reproducibility, our codebase and data processing scripts are available publicly [9].

Our main contributions include: (1) Collection and analysis of a large REN traffic dataset from Internet2 core routers, revealing cyclical patterns suitable for forecasting. (2) Design of a hybrid GRU-LSTM model tailored for REN traffic, capturing both long and short-term dependencies, and outperforming Prophet [10]. (3) Demonstration of prediction *lead time* as a critical hyperparameter in REN traffic forecasting. (4) Proposal to integrate anomaly detection to enhance prediction robustness.

The rest of the paper is organized as follows. Section II presents details about the dataset collected from the Internet2, and then presents some key insights from our preliminary analysis of the traffic. Section III presents the design of our traffic forecasting approaches. In particular, we justify the design choices made to achieve the desired outcomes and present model architectures. Section IV presents the results of our evaluation and discusses their consequences. Section V provides a review of related works in network traffic prediction. Section VI concludes our work and discusses the future directions.

II. Exploratory Data Analysis

In this section, we first present the details of the dataset collected from the Internet2 backbone. Following that, we present key insights gained from analyzing the traffic traces such as trends, cycles and other peculiar aspects that may potentially be related to the experimental workflows.

A. Dataset Collection and Overview

We collaborated with the Internet2 consortium [8] and collected NetFlow [11] traffic records from several core routers on the Internet2 optical backbone. Over a period of three months in late 2021, we recorded one in 1000 samples at each of the selected core routers on the Internet2 backbone, which resulted in a number of traffic records ranging from a few *millions* for less busy routers to *billions* for busier routers. In this paper, we primarily analyzed the traffic spanning the first month of our data collection period for two routers, one located in Atlanta, Georgia, and the other in Dallas, Texas. In its *GZip* compressed form, the Atlanta router dataset comprised of traffic records amounting to 8.49 GB, while the Dallas router amounted to 47.76 GB i.e. almost six times the amount of data for the Atlanta router.

The features of each record in the collected dataset are as follows. Features in bold were directly used for the analysis shown in this paper. (1) *export_sysid*: Internal unique identifier. **(2) *t_first*: Start time of the flow. (3) *t_last*: End time of the flow.** (4) *proto*: protocol number (not used). **(5) *src4_addr*: Source IPv4 address. (6) *dst4_addr*: Destination IPv4 address. (7) *src_port*: Source port number. (8) *dst_port*: Destination port number.** (9) *src_tos*: Type of service for the incoming interface. (10) *dst_tos*: Type of Service for the outgoing interface. **(11) *in_packets*: Number of incoming packets for this flow.** (12) *input_snmp*: Input SNMP index. (13) *output_snmp*: Output SNMP index. (14) *src_as*: Source BGP autonomous system number. (15) *dst_as*: Destination BGP autonomous system number. (16) *src_mask*: Number of bits in the source subnet mask. (17) *dst_mask*: Number of bits in the destination subnet mask. (18) *ip4_next_hop*: IPv4 address of the next hop router. **(19) *router*: name of the router.**

For each NetFlow record, we extracted the following attributes: *source and destination IP addresses, port numbers, packet counts, and protocol types*. During our initial data collection, each day's worth of data was stored as a *GZipped* file, resulting in 30 files per router. The dataset used in this paper includes *4.7 billion packets* and *97,539 flows* for the Atlanta router, and *56 billion packets* and *1,095,968 flows* for the Dallas router. To enhance processing efficiency, we performed the following preprocessing steps: (1) Extracted the required attributes from the raw NetFlow data. (2) Transformed the data into *Parquet* format with compression enabled, creating one Parquet file per router. This significantly improved data loading and processing speeds using Pandas. (3) Filtered out occasional missing data records caused by network timeouts or router maintenance. The complete dataset processing pipeline, including scripts and detailed instructions, is available in our GitHub repository.

B. Commercial Internet Traffic vs. REN Traffic

We performed a preliminary comparison of REN traffic patterns with those of public WANs. Specifically, we collected network layer traffic traces from a local Internet Service Provider (ISP) based in a mid-sized city in the continental US.

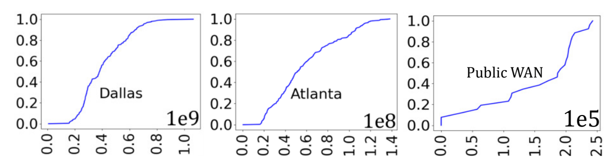

Fig. 2. CDF of the ***#Packets (incoming)*** (X-axis) among all the flows for the REN routers of Atlanta and Dallas, and a core router of an ISP (public WAN) in a mid-sized city in the continental US.

Figure 2 presents the *Cumulative Distribution Function (CDF)* of the ***#Packets*** (incoming) at the REN backbone routers in Atlanta and Dallas, and a core router from a commercial Internet Service Provider (ISP) for comparison. We observe that while the *median* value of ***#Packets*** is similar across REN and ISP routers, the distribution of ***#Packets*** for the REN routers is more uniform and less skewed compared to the ISP router. This uniformity stems from the consistent, sequential data flows typical in RENs, where large-scale scientific experiments drive traffic patterns. These flows generally involve large, consistent data transfers, unlike the more stochastic patterns in public WANs. This key observation allows us to effectively fine-tune the prediction *lead time* as a hyperparameter. The predictability of REN traffic flows suggests that an optimal *lead time* exists, which may not be as reliable in the more volatile traffic patterns of public WANs. Therefore, our approach specifically leverages the inherent predictability of REN traffic patterns to achieve high prediction accuracy with an appropriate choice of prediction *lead times*.

C. Traffic Analysis and Key Insights

Due to the volume of the traffic and the size of the dataset, we performed an initial exploratory analysis of the traffic to discover some key aspects of the communication patterns. Below, we briefly discuss our findings on this analysis.

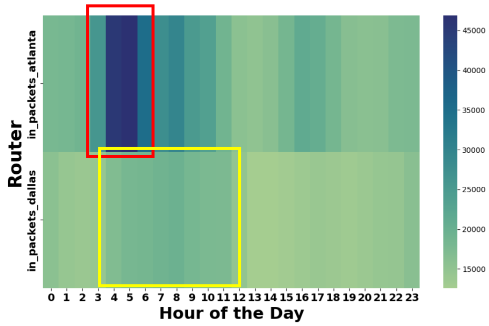

Fig. 3. Hourly Traffic Distribution (Average Traffic per Hour)

1) Diurnal Traffic Patterns: Our first observation is the presence of a marked diurnal increase in the traffic volume at both routers, albeit in different volumes and time periods as shown in Fig. 3. Atlanta experiences a high peak in traffic early in the day for the time period of 4-5 am (red square in Fig. 3), while Dallas sees a slight peak for the time period of 4 am-12 pm (yellow square in Fig. 3).

Figure 4 shows that the diurnal traffic patterns repeat every day for the whole duration of the dataset analyzed i.e. over 30 days. Another interesting finding is that weekdays corresponded to higher volumes in contrast to weekends, as expected from the schedules followed by academic and research institutions.

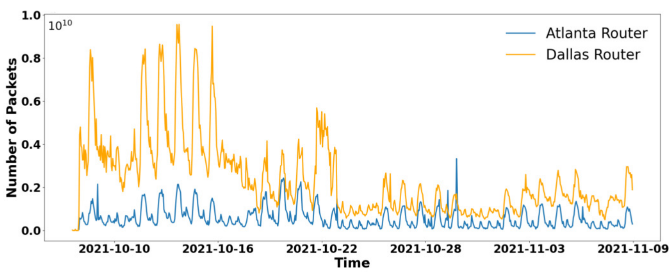

Fig. 4. Hourly traffic volumes of Atlanta and Dallas routers

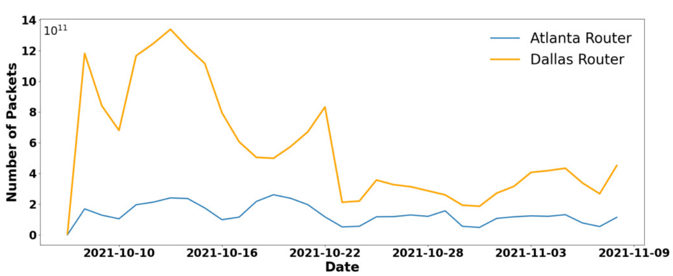

Fig. 5. Summarized daily volumes of traffic

2) Comparison of Traffic Volume: Atlanta vs. Dallas: Figure 4 shows the overall absolute traffic volume measured in packets for both routers, aggregated for every hour over 30 days. Figure 5 shows the daily aggregated traffic volumes for both routers. Figure 4 shows that Dallas experienced a sharp increase in daily traffic peaks from 2021-10-10 through 2021-10-15, followed by relatively lower volumes except for occasional anomalous peaks (e.g., around 2021-10-22). Atlanta experienced lower and uniform traffic volumes (and peaks) across all 30 days.

III. Traffic Forecasting for RENs

In this section, we present design details of our GRU-LSTM model for REN traffic prediction (see Sec. III-A) and then we present the comparison benchmark, namely Prophet which is a well-known time series prediction model from Facebook [10] (see Sec. III-B).

A. GRU-LSTM Hybrid Model

Long Short-Term Memory (LSTM) [12] networks are well-suited for time series forecasting due to their ability to capture long-term dependencies, while Gated Recurrent Units (GRUs) [13] efficiently handle short-term variations. Our hybrid GRU-LSTM model leverages these complementary strengths for improved accuracy in REN traffic prediction. This architecture is designed to overcome limitations of traditional RNNs by effectively managing vanishing gradients and capturing intricate temporal patterns in network traffic. The choice of a hybrid model is further supported by recent literature, which indicates that combined RNN architectures can outperform standalone models in complex forecasting tasks [14].

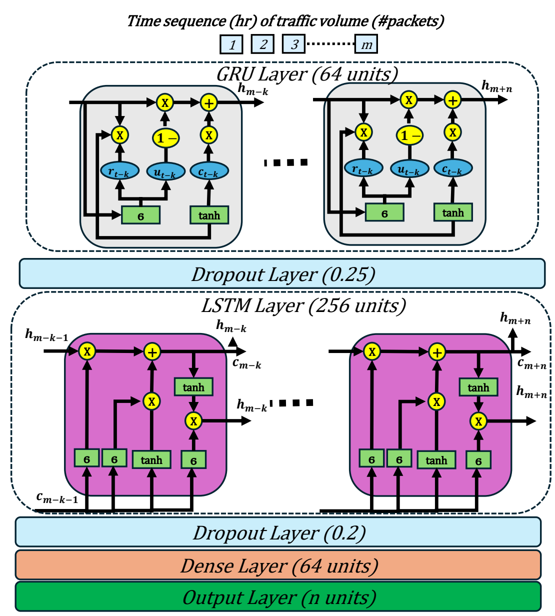

Fig. 6. Architecture of the Hybrid GRU-LSTM Model. The GRU-LSTM model is a *seq2seq* model, where the entire sequence output of *GRU* is passed down to the layers below it, including the *LSTM*. n is the *lead time* i.e. the number of time steps the model predicts for each iteration. m is the size of the input sequence that the model trains on to predict n steps into the future. The value of m was set to 64 i.e. input to the first layer of our GRU-LSTM model is a sequence of length 64.

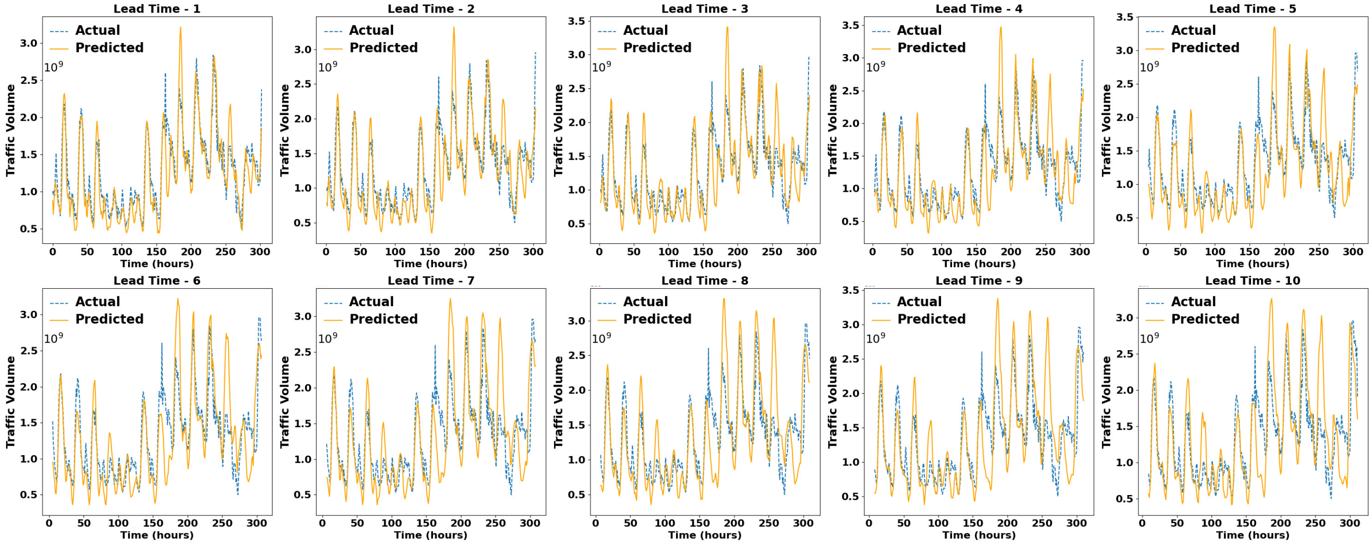

Fig. 7. Traffic Volume Prediction using our GRU-LSTM over 300 hours with *lead times* from 1 through 10.

1) Hyperparameter Tuning: We employed a systematic approach to hyperparameter tuning to optimize the performance of our GRU-LSTM model. We selected Adam as the optimizer due to its adaptive learning rate capabilities, which are beneficial for complex neural network training. Mean Squared Error (MSE) was chosen as the loss function to align with our objective of minimizing prediction errors in traffic volume. Evaluation metrics included MSE, MAE, R-squared (R^2), and MAPE to comprehensively assess the model's predictive accuracy and goodness of fit. The input sequence length (m) was set to 64 hours to capture sufficient historical context, and the prediction lead time (n) was varied from 1 to 10 hours to analyze its impact on forecasting accuracy. Further details on batch size, number of epochs, and dropout rates are available in the code repository.

B. Prophet

Prophet [10] is a forecasting approach designed for time series data using a decomposable model. It is particularly effective in learning typical patterns within time series, even when outliers or missing values are present, while simultaneously capturing seasonality and trend components. This capability makes it especially valuable for datasets that exhibit strong seasonal patterns, such as the daily and weekly cycles observed in Internet2 traffic. Moreover, Prophet is effective in capturing non-linear growth trends, such as those typical in RENs. The main features of Prophet are its ability to capture non-linear growth trends. We trained Prophet on a day-to-day traffic volume while choosing the hyperparameters that produced the best forecasting performance.

IV. Results and Discussion

In this section, we present the results of our GRU-LSTM REN traffic prediction model. First in Sec. IV-A, we present the influence of *lead times* on the model performance and also compare our GRU-LSTM model with a well-known traffic forecasting model, Prophet from Facebook [10] in Sec. IV-B.

A. GRU-LSTM Model Performance

Figure 7 presents the prediction results of our GRU-LSTM model over a 300-hour period for various *lead times*, ranging from 1 to 10 hours. As expected, the model's performance is highest at a *lead time* of 1, showing the lowest error metrics.

1) Impact of Lead Times: Lead time significantly influences the performance of our traffic prediction model. As shown in Figure 7, shorter lead times (e.g., 1-hour) yield higher accuracy but offer limited practical utility for network management. Predicting just one hour ahead provides minimal time for proactive adjustments in network configurations. Conversely, longer lead times, while providing more anticipation, inherently reduce prediction accuracy due to the increasing uncertainty in future traffic patterns.

The following are the *averaged* performance metrics of our GRU-LSTM model for all lead times tested:

- Mean Squared Error (MSE): $4.07e + 16$
- Mean Absolute Error (MAE): $153, 919, 069.77$
- R-squared: 0.868
- Adjusted R-squared: 0.867
- Mean Squared Log Error (MSLE): 0.0279
- Mean Absolute Percentage Error (MAPE): 0.1245

B. Prophet Model Performance

To provide a comparative analysis, we evaluated our GRU-LSTM model against Prophet, a widely-used time-series forecasting model. Figures 8 and 9 present the prediction results of Prophet for the Atlanta and Dallas routers, respectively. As shown in the Table I, Prophet's performance is significantly inferior to our GRU-LSTM model across all evaluation metrics. Although we only present the Prophet results for a *lead time of 1*, it is important to note that larger *lead times* resulted in even poorer performance (results omitted for brevity).

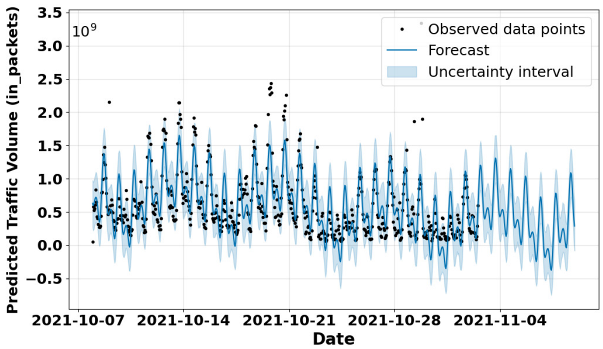

Fig. 8. Prophet Forecast for Atlanta router

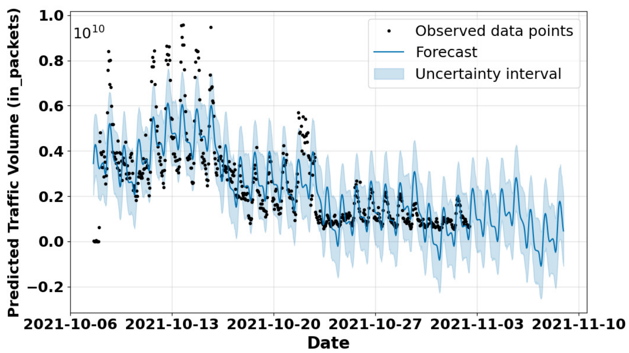

Fig. 9. Prophet Forecast for Dallas router

TABLE I
ROUTER EVALUATION METRICS COMPARISON FOR PROPHET

Metric	Atlanta Router	Dallas Router
MSE	5.137×10^{16}	8.375×10^{17}
MAE	180,122,812.23	772,377,811.41
R-squared	0.585	-2.152
Adjusted R-squared	0.582	-2.173
MAPE	0.760	0.580

While Prophet is effective in capturing seasonal patterns, it struggles to predict the sudden fluctuations and short-term changes in traffic that our GRU-LSTM model effectively captures. This difference underscores the advantage of using a hybrid RNN model that can learn from both long-term and short-term traffic characteristics, unlike the more generalized approach taken by Prophet. Our results clearly demonstrate the superior suitability of our approach for capturing the unique temporal dynamics of REN traffic.

C. Preliminary Traffic Anomaly Detection Approaches

To further enhance the performance of REN traffic prediction, we explored integrating anomaly detection methods. Our definition of a traffic anomaly is an unexpected surge or drop in traffic volume. Detecting such anomalies can improve traffic prediction by using anomaly information as an input feature to our GRU-LSTM model. Although the integration of anomaly detection with our traffic prediction model is reserved for future work, we present preliminary anomaly detection results using unsupervised learning approaches: Isolation Forest (IF) [15] and Local Outlier Factor (LOF) [16]. We applied both methods to the Atlanta and Dallas router traffic data. Figure 10 and Figure 11 show the detected anomalies using IF and LOF, respectively. The *contamination* parameter was set to 0.05 for both methods, and the number of neighbors for LOF was set to 20. Both IF and LOF effectively identify anomalies, marked in Figures 10 and 11. These anomalies often correspond to sudden traffic spikes or drops following peak traffic periods, indicating the potential for anomaly detection to enhance traffic prediction accuracy. Future research will focus on experimenting with different contamination values and incorporating anomaly detection as a feature in the GRU-LSTM model.

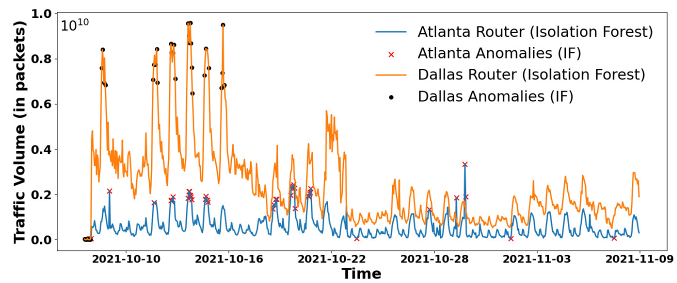

Fig. 10. Anomaly Detection Using Isolation Forest (IF)

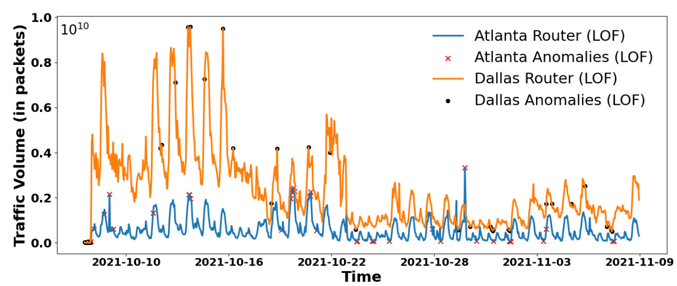

Fig. 11. Anomaly Detection Using Local Outlier Factor (LOF)

V. RELATED WORK

Network traffic prediction is a well-explored domain, crucial for efficient network management and resource allocation. Various techniques, from traditional statistical models to advanced machine learning, have been used to forecast traffic in a variety of network types. Early studies, like Vujicic et al. [17], used statistical methods for traffic prediction in public safety networks. Ramakrishnan and Soni [18] later demonstrated RNNs to capture temporal dynamics in network traffic, paving the way for deep learning applications. More recently, deep learning, including hybrid models, has shown promise in boosting prediction accuracy, shifting from traditional methods to more complex architectures that can capture nonlinearities and improve forecasting precision.

Ferreira et al. [19] highlighted the importance of RNNs, especially LSTMs, and GRUs for their superior forecast performance for network traffic over traditional methods. However, they note the need to explore beyond standard RNNs and similar deep learning-based methods to further improve forecast performance. Specifically, they advocate for models exceeding *"now rather standard RNNs, CNNs or GNNs"* to improve accuracy, directly supporting the hybrid GRU-LSTM model presented in this paper. This model combines GRUs for their short-term dependency modeling and LSTMs for their long-term dependency modeling, two features prevalent in REN traffic. Although models such as the Traffic Transformer [20] and Diffusion Convolutional Recurrent Neural Networks [1] exist, their effectiveness in REN traffic forecasting, with its specific characteristics, is underexplored. Addressing these gaps, our GRU-LSTM model, evaluated on real REN data with *lead time optimization*, bridges the gap for REN traffic prediction.

VI. Conclusion & Future Work

This paper introduces a novel traffic prediction approach specifically designed for Research and Education Networks (RENs), addressing the gap in existing traffic forecasting techniques that are primarily tailored for public WANs. We presented a comprehensive analysis of real-world traffic data from the Internet2 backbone, revealing unique traffic patterns and cyclical behaviors inherent to RENs. Our key contribution is the development of a hybrid GRU-LSTM model, which effectively captures both short-term and long-term dependencies in REN traffic, outperforming the benchmark Prophet model. We demonstrated the critical impact of prediction *lead time* as a hyperparameter, showing that optimal lead times can be identified to balance prediction accuracy and practical network management needs. Furthermore, we explored preliminary traffic anomaly detection methods, suggesting a pathway to enhance prediction models by integrating anomaly detection capabilities.

Future work will focus on several directions: (i) extensive hyperparameter fine-tuning to further optimize model performance, (ii) expanding our analysis to a broader range of core routers and longer data collection periods within the Internet2 backbone, (iii) in-depth investigation of the impact of varying *lead times* on model accuracy and practical network operations, and (iv) ensembling anomaly detection mechanisms with our GRU-LSTM model to create a more robust and adaptive traffic prediction system. These future enhancements aim to further validate and improve our approach, providing valuable tools for network management and resource optimization in critical REN infrastructures.

VII. Acknowledgments

We would like to thank the Internet2 consortium for their support without which this work would not have been possible. This work is sponsored in part by National Science Foundation (NSF) grant with award number OAC-2322369.

References

[1] D. Andreoletti, S. Troia, F. Musumeci, S. Giordano, G. Maier, and M. Tornatore, "Network Traffic Prediction based on Diffusion Convolutional Recurrent Neural Networks," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 246–251.

[2] A. Saha, N. Ganguly, S. Chakraborty, and A. De, "Learning Network Traffic Dynamics Using Temporal Point Process," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1927–1935.

[3] L. Nie, Z. Ning, M. S. Obaidat, B. Sadoun, H. Wang, S. Li, L. Guo, and G. Wang, "A Reinforcement Learning-Based Network Traffic Prediction Mechanism in Intelligent Internet of Things," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 3, pp. 2169–2180, 2021.

[4] D. Nadig, B. Ramamurthy, B. Bockelman, and D. Swanson, "APRIL: An Application-Aware, Predictive and Intelligent Load Balancing Solution for Data-Intensive Science," in *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 2019, pp. 1909–1917.

[5] A. Bhattacharjee, A. D. Chhokra, Z. Kang, H. Sun, A. Gokhale, and G. Karsai, "BARISTA: Efficient and Scalable Serverless Serving System for Deep Learning Prediction Services," in *2019 IEEE International Conference on Cloud Engineering (IC2E)*, 2019, pp. 23–33.

[6] V. S. S. L. Karanam and B. Ramamurthy, "Collective deep network matrix factorization for large-scale modeling of the internet backbones," in *2023 IEEE Future Networks World Forum (FNWF)*. IEEE, 2023, pp. 1–6.

[7] K. Begum, M. M. Rashid, and M. A. U. Shariff, "Comparative Study of Big Data Visualization Tools and Techniques," in *Applied Informatics for Industry 4.0*. Chapman and Hall/CRC, 2023, pp. 188–199.

[8] F. Yeung, "Internet 2: Scaling Up the Backbone for R&D," *IEEE Internet Computing*, vol. 1, no. 2, pp. 36–37, 1997.

[9] M. A. U. Shariff, V. S. S. L. Karanam, and B. Ramamurthy, "REN-Traffic-Prediction-GRU-LSTM," https://github.com/mhnarfth/REN-Traffic-Prediction-GRU-LSTM, 2024.

[10] S. J. Taylor and B. Letham, "Forecasting at scale," *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.

[11] B. Claise, "Cisco systems netflow services export version 9," Cisco Systems Inc., Tech. Rep., 2004.

[12] S. Hochreiter, "Long Short-term Memory," *Neural Computation MIT-Press*, 1997.

[13] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[14] E. Hossain, M. A. U. Shariff, M. S. Hossain, and K. Andersson, "A novel deep learning approach to predict air quality index," in *Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020*. Springer, 2020, pp. 367–381.

[15] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*. IEEE, 2008, pp. 413–422.

[16] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.

[17] B. Vujicic, H. Chen, and L. Trajkovic, "Prediction of traffic in a public safety network," in *2006 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2006, pp. 4–pp.

[18] N. Ramakrishnan and T. Soni, "Network traffic prediction using recurrent neural networks," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018, pp. 187–193.

[19] G. O. Ferreira, C. Ravazzi, F. Dabbene, G. C. Calafiore, and M. Fiore, "Forecasting Network Traffic: A Survey and Tutorial With Open-Source Comparative Evaluation," *IEEE Access*, vol. 11, pp. 6018–6044, 2023.

[20] L. Cai, K. Janowicz, G. Mai, B. Yan, and R. Zhu, "Traffic transformer: Capturing the continuity and periodicity of time series for traffic forecasting," *Transactions in GIS*, vol. 24, no. 3, pp. 736–755, 2020.