# Dynamic Traffic Prediction Model Retraining for Autonomous Network Operation

**Fatemeh Tabatabaeimehr[1], Luis Velasco[2], Marc Ruiz[2], Hamzeh Khalili[1], Ramon Aparicio-Pardo[3]**
*[1] Centre Tecnològic de Telecomunicacions de Catalunya (CTTC/CERCA), Castelldefels, Spain*
*[2] Optical Communications Group (GCO), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*
*[3] Université Côte d'Azur, CNRS, Inria, I3S, France*
*e-mail: ftabatabaeimehr@cttc.es*

**ABSTRACT**
In general, the availability of an accurate machine learning (ML) model plays a particularly important role in the development of new networking solutions and is one of the main drivers for the development of 5G and beyond networking. Although an option is to update the model once inaccurate data is detected, such approach requires high computational effort, specially once the data history is large. In this paper, we propose an approach that combines a traffic prediction model based on Long Short-Term Memory (LSTM) with an analysis module for dynamic connection capacity allocation. Once the model is generated, re-training can be triggered after inaccuracies are detected by the analysis module. Illustrative numerical results show the benefits from the proposed decision-based re-training approach to reduce the number of re-training rounds while maintaining model accuracy.
**Keywords**: autonomous operation, metro network, traffic prediction, model retraining, LSTM

## 1. INTRODUCTION

The emergence of 5G technology has revolutionized the telecommunications industry [1], necessitating significant changes not only in the way applications are designed but also in their management [2]. In particular, the increasing use of machine learning (ML)-based applications that convey traffic flows with unknown characteristics [3] presents a significant challenge in terms of operational goals such as guaranteeing quality of service (QoS) without incurring in an excessive capacity overprovisioning [4]. In this regard, the traffic should be properly forecasted, which poses a challenge since some traffic characteristics can be complex and hard to model due to the presence of multiple periodicities ranging from a few hours to several days. This fact makes it impractical applying traditional predictive approaches in many multilayer optical network automation scenarios, e.g., dynamically allocating capacity to a traffic flow according to traffic prediction [5].

To address this challenge, various machine learning techniques have been proposed to predict time series data accurately [6]. Long Short-Term Memory (LSTM) networks have been proposed as a means of expanding temporal dependence learning [3]. However, despite the high overall accuracy of LSTM models, they are still subject to errors when pre-trained, which can reduce their robustness in autonomous network operation. Under-predicting traffic can lead to connection capacity under-provisioning and result in traffic loss.

In this paper, we propose an approach that combines an LSTM-based traffic prediction model with a *Model Refining Analysis* module tailored for solving the dynamic connection capacity allocation problem. Our approach aims to reduce the need for re-training models while maintaining accuracy. By implementing this approach, we expect to enhance the reliability and efficiency of autonomous network operation, especially in multilayer optical network automation scenarios. The paper is organized as follows: Section 2 introduces the procedure of decision-based LSTM model re-training with real-time error detection, while Section 3 details the models and procedures. Section 4 discusses performance evaluation and results, and Section 5 concludes the paper.

## 2. DECISION-BASED MODELS RE-TRAINING

In this section, we propose an autonomous capacity management system for connections transporting packet traffic flows using continuous re-training. Our approach considers a connection to be either a customer connection transporting a few Gb/s traffic flow or a virtual link supporting flow aggregation with traffic of hundreds Gb/s. The objective is to allocate the minimum capacity required to support the flow for the next period while meeting the intended performance, such as avoiding loss due to capacity under-provisioning.

Figure 1 illustrates the two basic approaches for autonomous capacity management: i) threshold-based (Figure 1a); ii) pre-trained LSTM-based prediction (Figure 1b). Figure 2 sketches the expected evolution of the allocated capacity (solid-colored lines) for a traffic flow (dotted line) that experiences a high perturbation in traffic. Under the threshold-based approach, future capacity is reactively adjusted according to the current traffic. A threshold value is statically defined, which must be set low enough to guarantee no loss during sharp traffic changes, which leads to high overprovisioning. In contrast, a LSTM model that be configured offline is used for dynamic operation. Notwithstanding the good overall accuracy, LSTM underestimates traffic during the peak, which can lead to loss, and produces capacity under-provisioning. In this regard, we propose decision-based approach, a combination of LSTM-based traffic prediction model and Model Refining Analysis module for capacity allocation (Figure 1c). The Model Refining Analysis module includes a *model drift detection algorithm* (MDDA) that integrates with LSTM-based traffic predictor to compute: *i*) prediction error, and *ii*) check whether the error is out of a specific boundary. In this approach, the LSTM-based model can efficiently learn the new pattern in the traffic without requiring continuous re-training for each individual data point. The formal details of the extended environment blocks are presented in the next section.

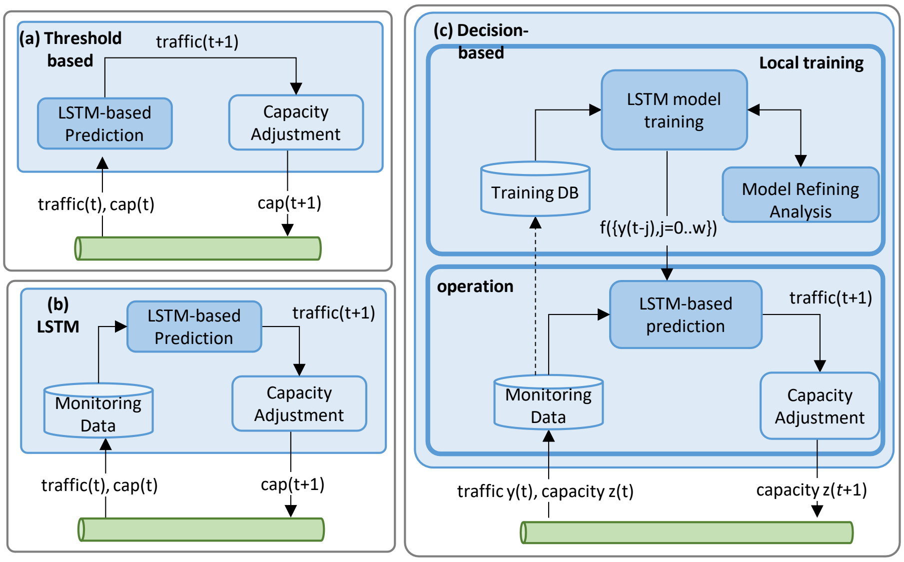


Figure 1 (a) Threshold-based, (c) LSTM, and (c) Decision-based.

## 3. MODELS AND PROCEDURE

We consider that a monitoring data set is available for every connection, which consists of samples <y(t), z(t)>, with the traffic and capacity at time *t*, respectively. For the sake of simplicity, let us consider that a LSTM model is represented by eq. (1) has been previously trained on a dataset of historical network traffic data (*tunedLSTM* model). That historical data consists of one week of data points with the last *w* traffic monitoring samples and produces traffic predictions $\hat{y}$ with single upper confidence intervals to be used at time *t*.

As the initial step, model continuously trained in an interval [0, t] and gaussian mean square error (gssMSE) of actual traffic y(t) and traffic prediction y^(t) for interval [0, t] is calculated (eq. (2)). As a result, it returns the normal distribution of MSE characterized with a range of standard deviation α (error tolerance) (eq. 3). To increase the amount of data available for analysis within a α range, we employ a random variable θ to improve the overall accuracy of our results. Moreover, Model Refining Analysis module should select the most appropriate range of threshold from the range of standard deviation to keep the performance as continual re-training while the number re-training reduces as much as possible.

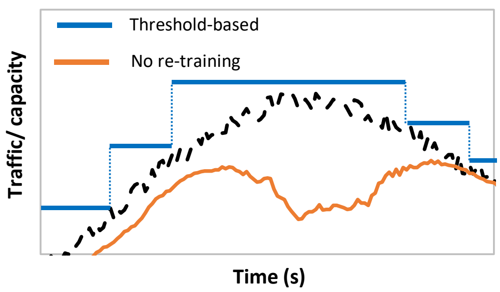


Figure 2 Expected capacity allocation.

$$tunedLSTM \sim f(\{y(t-j), j=0..w\}) \tag{1}$$

$$gssMSE \sim \varphi\big(mse\big(y(1 \dots t), \hat{y}(1 \dots t)\big), \alpha^2\big) \tag{2}$$

$$gssMSE = [\alpha, \dots, m\alpha \pm \theta] \quad m \in \mathbb{Z}^+,\ \theta \in \mathbb{R} \tag{3}$$

$$targetErr = \max\big(\omega(gssMSE)\big)/n \tag{4}$$

Clearly, relaxing the error tolerance has impact on the model's performance, with an increase in prediction error and a decrease in number of model re-training. This highlights the importance of carefully selecting and tuning input variables for accurate and reliable predictions. To this aim, the targetERR (eq. (4)) is computed as the mean of maximum frequency distribution of the gssMSE error. Once the error is above the targetErr, algorithm stops searching the rest of the list. Once the error tolerance is fixed, tuneLSTM model is ready for decision-based re-training. Note that the selected error tolerance variable can be updated with a given periodicity. The algorithm assesses the performance of the model and determines whether to re-train the model with an unknown traffic data sample.

## 4. ILLUSTRATIVE RESULTS

For the purpose of evaluating our method, we generated synthetic traffic data for two traffic flows over a period of three months. The traffic data includes a sinusoidal traffic component with two different fluctuation variables, 0.01 and 0.25. In this paper, we evaluated our approach using traffic data with a daily profile pattern. We used the first half of the traffic data for training and the second half for validating the proposed approach. After evaluating a wide range of the number of hidden layers and size, we selected a fully-connected LSTM-model with three hidden layers and 30, 40, and 20 neurons per layer, respectively. We used the backpropagation training algorithm with batches of 128 samples and epoch 50. As the initial evaluation let us first see how we determine the proper error threshold to configure the analysis module.

Figure 3a and b shows the tradeoff between prediction error and number of re-training rounds with respect to the error tolerance for traffic flow with fluctuation 0.01 and 0.25. As the error tolerance increases, the number of re-training decreases until it reaches a target error, $\sim 3\alpha \pm \theta$ determined according to our proposed approach; at which point the searching loop terminates. Let us assume that the traffic monitoring data is received with a frequency of 1 per second. Figure 3c and d shows the behavior of all the approaches at some intervals of the simulation. In no update approach, model is not able to anticipate enough the traffic changes, which leads to capacity under-provisioning and traffic loss, whereas the continual re-training and decision-based re-training approaches fit capacity remarkably close to actual traffic, resulting into a very good performance in terms of capacity provisioning. Table 1 also summarizes the numerical simulation results, where model does not re-train, continuously re-train, and re-trained based on decision from analysis module. Decision-based produces slightly less number of re-training compare to continual re-training.

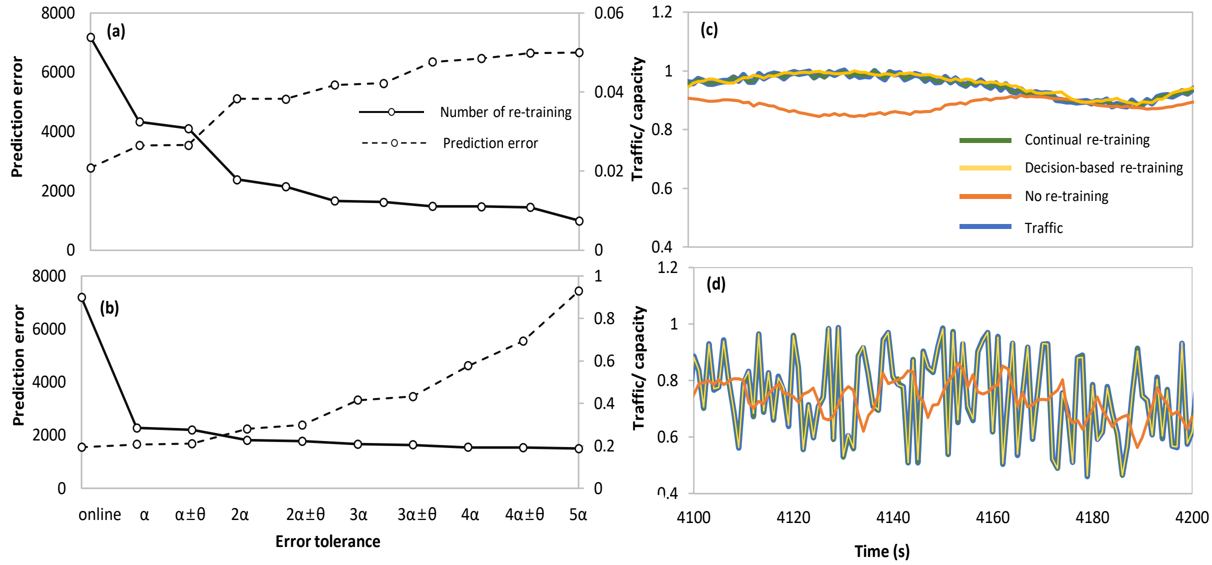


Figure 3. Approaches comparison for 0.1 traffic (a, c) and 0.25 (b, d) traffic.

Table 1 Model performance summary

| Approach | Mean square error | | Number of re-training | |
|---|---|---|---|---|
| | **0.01** | **0.025** | **0.01** | **0.025** |
| Continual re-training | 0.000002 | 0.004143 | 7200 | 7200 |
| Decision-based re-training | 0.00005 | 0.005035 | 2385 | 2635 |
| No re-training | 0.001335 | 0.123427 | - | - |

After fixing the error threshold, we configure the *Model Refining Analysis* module. Figure 4a and b presence the error convergence over several days in continual re-training and decision-based re-training approaches. We observe that decision-based approach converges to the minimum error like continual approach.

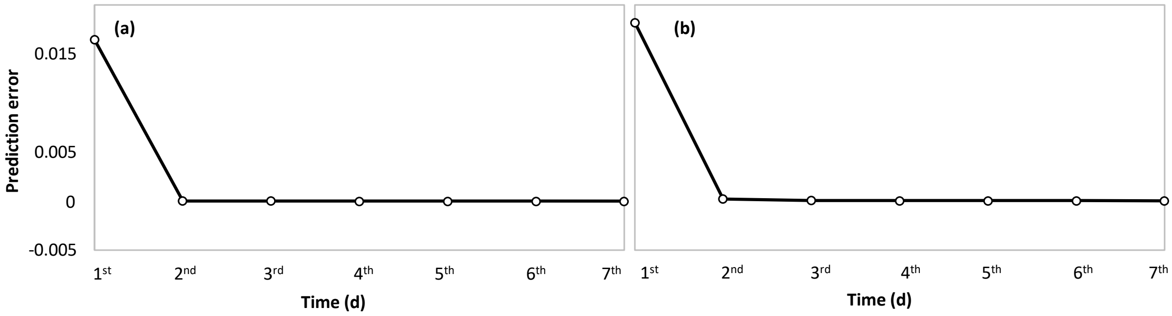


Figure 4. Accuracy performance result.

## 5. CONCLUSIONS

In this paper, we proposed an autonomous connection capacity management approach based on an analytic module for long-term traffic prediction. Our approach deploys highly accurate LSTM models and overcomes critical obstacles such as the need for pre-trained models with old historical measurements. Our proposed method reduces the number of re-trainings, into ~33% and ~36%, required while maintaining high model performance, resulting in moderate capacity overprovision. Overall, our results show that both the continual re-training and decision-based re-training approaches require slightly lower over-provisioning than the no re-training approach. This suggests that our approach is effective in managing connection capacity while maintaining accuracy. Future work may involve extending our approach to consider multiple traffic flows and dynamically adjusting connection capacity based on real-time traffic information.

## ACKNOWLEDGEMENTS

This work was partially funded by MCIN/AEI/ 10.13039/501100011033 grant PID2021-126431OB-I00 (ANEMONE), and Generalitat de Catalunya grant 2021 SGR 00770. This work also was funded by the European Commission Horizon Europe SNS JU DESIRE6G project (G.A. 101096466), from the Spanish MINECO and the European Union though the NextGeneration UNICO5G TIMING (TSI-063000-2021-145), and the AEI IBON (PID2020-114135RB-I00) projects. The authors also acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-19-CE-25-0001-01 (ARTIC project).

## REFERENCES

[1] Misra, S., & Bansal, S. (2019). 5G: Architecture, challenges, and future scope. 2019 5th International Conference on Computing Communication and Automation (ICCCA), 1-6.

[2] L. Velasco, S. Barzegar, F. Tabatabaeimehr, and M. Ruiz, "Intent-Based Networking for Optical Networks [Invited Tutorial]," IEEE/OPTICA Journal of Optical Communications and Networking (JOCN), vol. 14, pp. A11-A22, 2022.

[3] F. Tabatabaeimehr, S. Barzegar, M. Ruiz and L. Velasco, "Combining Long-Short Term Memory and Reinforcement Learning for Improved Autonomous Network Operation," 2021 Optical Fiber Communications Conference and Exhibition (OFC), San Francisco, CA, USA, 2021, pp. 1-3.

[4] S. Barzegar, M. Ruiz, and L. Velasco, "Packet Flow Capacity Autonomous Operation based on Reinforcement Learning," MDPI Sensors, vol. 21, pp. 8306, 2021.

[5] F. Morales, Ll. Gifre, F. Paolucci, M. Ruiz, F. Cugini, P. Castoldi, and L. Velasco, "Dynamic Core VNT Adaptability based on Predictive Metro-Flow Traffic Models," IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 9, pp. 1202-1211, 2017.

[6] Wang, K., Lin, Y., Wang, G., Zhang, Y., & Chen, Y. (2021). Network traffic prediction using long short-term memory neural networks. In Proceedings of the 2021 International Conference on Computer Science, Communications and Networks (CSCN 2021) (pp. 373-379).