NWDAF Implementation: Network Function Load Analytics in 5G Core Network

Quang-Vinh Tran*, Tuan-Bach Phan†, Quang-Diep Pham‡
Viettel High Technology Industries Corporation, Viettel Group
Hanoi, Vietnam
Email: {*vinhtq24,†bachpt1,‡dieppq}@viettel.com.vn

***Abstract*—Multivariate data is prevalent in telecom systems and requires efficient processing to extract actionable insights. Leveraging such data, we developed the NWDAF module with a focus on ”Network Function (NF) load analytics” to improve Viettel Group’s network performance. Using operational data from Viettel’s 4G Mobility Management Entity (MME) serving as a proxy for 5G behavior our framework supports real-world analysis and optimization of network functions in next-generation deployments.**

***Index Terms*—NWDAF, NF load, 5G network, NF selection, correlation.**

I. Introduction

In the 5G core network, NWDAF, as defined by 3GPP TS 23.288 (Release 18) [1], is a specialized module for collecting data from multiple sources and delivering analytics, such as NF load statistics and predictions, to network functions. Accordingly, Viettel’s 5G network implements the ”NF Load Analytics” service to fulfill this requirement and facilitate data-driven network function management.

Among the inputs for implementing this service, NF resource usage has the potential to provide immediate responsiveness. This parameter, collected from OAM - Operations, Administration and Maintenance (with the deployed network utilizing the MANO system) [2]. Resource data from servers within the K8s cluster, where network functions are deployed, will be gathered, with NWDAF responsible for performing analysis (statistical and/or predictive). The analysis of data to support value prediction will be the primary focus of this study.

With the ability to predict information about NFs in the core network, the NF load analytics process contributes significantly to supporting various procedures within the network. Below are some notable examples:

- NF Selection - Facilitates optimal resource allocation by enabling NF consumers to dynamically select service instances based on real-time load metrics and historical performance patterns
- Anomaly Detection and Prediction - AI models can identify irregular load patterns and predict potential service degradation before it impacts network performance by analyzing historical data trends and real-time network metrics

II. NF load data

3GPP Input. We focus on features that characterize the resource consumption and load levels of Network Functions (NFs). Prioritization of these features is based on the targeted use cases and the capability of NFs to provide such information. The analytical process is described in Fig. 1.

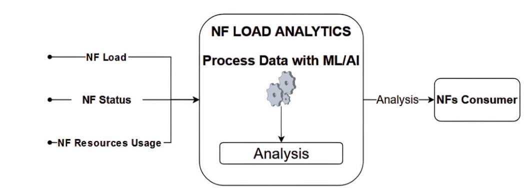

Fig. 1. NF Load analytics flow.

4G MME-dataset. The study analyzed a dataset of mobile network traffic traces collected over 7 days (March 4-10, 2024), representing normal network operating conditions without disruptions. The dataset encompasses key system performance metrics, for this research purposes, only CPU utilization ratio and is expressed as a percentage will be use.

$$x_{cpu} = \frac{Total\ CPU\ Resources}{CPU\ Resources\ Consumed} \times 100\% \quad (1)$$

Our network consists of multiple independent sites, each with 32 machines linked to core network functions; thus, each prediction model utilizes 32 features.

III. Analytic and Conclusion

In the context of multivariate data analysis, correlations can reveal underlying dependencies between features, which are crucial for understanding data structures and improving model design. The most commonly used metric to calculate correlation is Pearson’s correlation coefficient [3], defined mathematically as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (2)$$

where x_i and y_i are the data points, $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively, and n is the number of observations.

The correlation between CPUs is low, with significant correlations only observed in specific pairs such as $0-16$, $1-17$, etc., ranging from 0.3 to 0.45. Therefore, each feature holds distinct and meaningful information in the prediction process.

979-8-3315-5678-5/25/$31.00 ©2025 IEEE

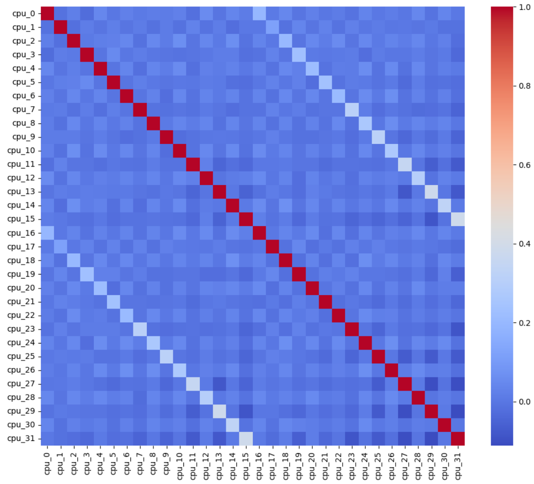

Fig. 2. Correlation matrix

Given time series data with cyclical characteristics, we utilize models that capture temporal dependencies for forecasting purposes. RNN models and their variants, including Transformer models, will be evaluated. Notably, the LSTM model [4], with its ability to retain information over longer time spans, has demonstrated superior performance on such data.

The training process is performed on overlapping sliding windows to augment the data, using training pairs defined as follows:

$$\begin{cases} X^{(i)} = [x_i, x_{i+1}, \dots, x_{i+k-1}] \in \mathbb{R}^{k \times d} \\ Y^{(i)} = [x_{i+k}, x_{i+k+1}, \dots, x_{i+k+m-1}] \in \mathbb{R}^{m \times d} \\ i = 1, 2, \dots, T - (k+m) + 1 \end{cases} \tag{3}$$

The inference pairs used for evaluation follow a similar approach, which helps reduce the cumulative error by predicting over each time window rather than on individual samples.

Let x_j denote the actual value and $\hat{x}_j$ the predicted value (for detailed calculations, see [4] and [5]), we define the following evaluation metrics for predictions:

$$\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |x_j - \hat{x}_j| \tag{4}$$

$$\text{MAPE} = \frac{1}{n} \sum_{j=1}^{n} \left| \frac{x_j - \hat{x}_j}{x_j} \right| \times 100 \tag{5}$$

We adopt a prediction window of 30 minutes ($m = 30$), for which a training window of 60 minutes ($k = 60$) yields the best performance. Baseline for result comparison: RNN, GRU, and Transformer. Tab. I shows the superior performance of LSTM (3-layer) compared to other models in terms of MAE (0.58) and MAPE (0.94) metrics.

The flow of the "NF load analytics service" is illustrated in Fig. 4, where data on the usage of NF resources are collected from OAM. The consumer of the NRF (Network Repository Function), upon receiving information about the desired NF instances, will request NWDAF to analyze and determine which NFs are the most available (for example, those that are not likely to be overloaded or are expected to have more free resources in the near future).

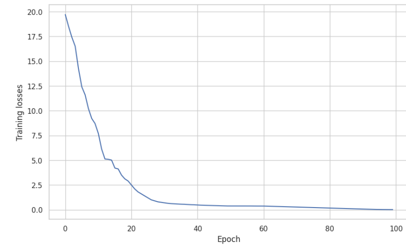

Fig. 3. Training phase of LSTM model.

TABLE I
MAPE AND MAE OF PREDCTION METHODS.

Metric	Site	Model			
		RNN	LSTM	GRU	TRANS
MAE	SG**01	1.02	0.47	0.65	1.29
	SG**02	2.12	0.61	0.74	0.51
	SG**03	1.79	0.33	0.51	0.27
	SG**04	2.24	0.90	1.10	0.89
	average	1.79	0.58	0.75	0.74
MAPE (%)	SG**01	2.45	1.12	1.56	3.10
	SG**02	3.15	0.90	1.10	0.76
	SG**03	2.90	0.53	0.83	0.44
	SG**04	3.02	1.22	1.84	1.21
	average	2.88	0.94	1.33	1.38

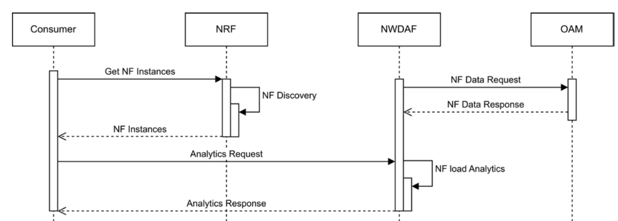

Fig. 4. NF load analytics flow.

With the predictive capabilities of LSTM and the extensive dataset provided by Viettel, we anticipate the practical feasibility of implementing NF load analytics. In the future, we plan to expand our experiments to include data on memory and storage resource utilization, as well as collect load data of NFs from the 5G core network.

REFERENCES

[1] 3rd Generation Partnership Project, "Architecture enhancements for 5g system (5gs) to support network data analytics services," *Telecommucation Management*, vol. 3GPP TS 23.288, 2024.

[2] ——, "Management and orchestration; 5g performance measurements," *Telecommucation Management*, vol. 3GPP TS 28.552, 2024.

[3] K. Pearson, "Vii. note on regression and inheritance in the case of two parents," *proceedings of the royal society of London*, vol. 58, no. 347-352, pp. 240–242, 1895.

[4] J. Schmidhuber, S. Hochreiter *et al.*, "Long short-term memory," *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.

[5] R. Ghanbari and K. Borna, "Multivariate time-series prediction using lstm neural networks," in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*. IEEE, 2021, pp. 1–5.