Network Slice Traffic Demand Prediction for Slice Mobility Management

Muhammad Ashar Tariq, Malik Muhammad Saad, Mahnoor Ajmal,
Ayesha Siddiqa, Junho Seo, Yang Haishan, Dongkyun Kim
School of Computer Science and Engineering, Kyungpook National University, Republic of Korea
{tariqashar, maliksaad, mahnoor.ajmal, asiddiqa, jhseo, yanghs158, dongkyun}@knu.ac.kr

***Abstract*—5G revolutionizes connectivity with network slicing, but non-uniform slice deployment poses challenges. Efficient slice handover mechanisms and predictive learning offer solutions. We propose Long Short-Term Memory (LSTM) based prediction of network slice traffic demands for proactive resource management to resolve challenges of mobility management and non-uniform slice deployment within a registration area. Thereby, allowing seamless service provision throughout the registration area. Our proposed model shows excellent results in predicting the number of received requests for all available slices. These results are then utilized to deduce the average slice demand for each service in the network. The obtained results are beneficial for configuring the slice resources proactively in order to provide seamless service to mobile users throughout the registration area, regardless of the non-uniformity in slice deployment among different tracking areas.**

***Index Terms*—5G, 5G-Advanced, Network Slicing, Mobility management, Slice Handover, Resource Management.**

I. Introduction

5G technology heralds a transformative era in connectivity, delivering unprecedented versatility and performance in wireless networks. Central to this revolution is network slicing, a pivotal architectural innovation exemplifying 5G's dynamic capabilities. Network slicing facilitates the creation of dedicated, isolated virtual networks within a shared physical infrastructure, tailored for diverse use cases and applications [1]. This paradigm shift hinges on the principle of resource customization to meet specific demands. However, amidst network slicing's promises of flexibility, a critical challenge surfaces: mobility management within these intricately partitioned network slices. Mobility, essential for seamless device and user movement across the network, becomes intricate when each slice operates as a distinct, self-contained ecosystem with unique performance criteria. The importance of addressing mobility management within network slicing cannot be overstated. As the 5G ecosystem expands and diversifies, mobile networks face escalating demands. Whether for autonomous vehicle navigation, immersive augmented reality, or the orchestration of massive IoT deployments, uninterrupted mobility is paramount [2]. It profoundly influences service quality, user experiences, and the seamless execution of mission-critical applications. Consequently, resolving mobility management challenges within network slicing is pivotal for unlocking 5G's full potential and harnessing its capabilities across an array of industries.

One of the predominant challenges inherent to 5G network slicing is the deployment of slices that are non-uniform across the network. The Third Generation Partnership Project (3GPP) assumed a homogeneous service support among all tracking areas within a registration area until Rel-17. However, in terms of real deployments, it is not feasible. 3GPP is now working on the challenge of non-uniform deployment of slices within the registration area in its Rel-18 [3]. This non-uniformity can lead to several adverse consequences, including service interruptions when the slice service area border is crossed as well as non-homogeneous service support throughout the registration area. In a dynamic 5G ecosystem, where various slices coexist, each catering to distinct service requirements, the proper orchestration and deployment of these slices become crucial. When not adequately managed, this non-uniformity can result in disparities in service quality and availability. This non-homogeneous service support can impede the realization of 5G's promise of ubiquitous, seamless connectivity and underscores the critical role that effective slice management plays in addressing these challenges and ensuring a consistent and reliable user experience across the entire registration area [4].

Addressing the challenge of non-uniform slice deployment is crucial but deploying new slices uniformly across the registration area can be cost-prohibitive. Hence, exploring cost-effective alternatives is essential. Efficient slice handover emerges as a promising solution, mitigating mobility management issues and ensuring seamless service delivery. Efficient slice handover mechanisms dynamically allocate network resources, eliminating the need for extensive uniform slice deployment. This adaptive approach optimizes resource utilization and minimizes service interruptions as users and devices move between different tracking areas within a registration area [5]. Furthermore, in anticipation of 3GPP's intelligence infusion into networking operations in Rel-18, integrating prediction learning into network management proves wise. Predictive models can foresee slice demand patterns, enabling proactive resource allocation for incoming traffic from various slices. It can also equip network operators to meet the evolving demands of the dynamic 5G landscape. representing a forward-thinking solution that harmonizes deployment challenges, operational efficiency, and the delivery of high Quality-of-Service (QoS) across the registration area. Hence, in this paper, we propose a long short-term memory (LSTM) based solution to predict network slice traffic demands that can be utilized for proactive resource management in target tracking areas considering user mobility. Our solution is able to predict the traffic demands of individual slices based on the previously received service requests. This can solve the issue of mobility management within a registration area by implementing efficient resource management among available network slices in order to provide seamless service to the users within the registration area, regardless of the availability of specific network slices.

The rest of the manuscript is organized as follows. Section II summarizes the literature regarding network slice mobility management, slice handover, and traffic demand prediction. Section III presents our proposed system model. Section IV describes the proposed scheme with its implementation and evaluation. Finally, in section VI, conclusions and future work are presented.

II. Related Work

The concept of network slicing has become crucial in adapting network services to meet the diverse demands of modern communication. However, the deployment of non-uniform slices and the dynamic nature of mobile networks present a range of challenges when it comes to maintaining seamless mobility for these slices. Managing the mobility of network slices involves intricate issues such as handovers, resource allocation, and QoS maintenance, which require innovative solutions to fully capitalize on the advantages of network slicing. While handovers are essential in mobile networks, they can also be complex in terms of QoS, resource consumption, and overall network performance.

The authors in [1] examine the challenges in network slicing mobility management, providing a thorough survey that not only identifies these problems but also offers insightful information on mobility trends, user grouping, and slice mobility triggers. Their research focuses on numerous aspects of slice mobility, such as optimal patterns, the Follow Me Edge idea, service migration in Infrastructure as a Service (IaaS) clouds, key triggers, and UE grouping approaches. Furthermore, the study provides a preliminary evaluation of enabling technologies such as system virtualization and SDN for resource allocation in network slices, providing a full field overview. This study greatly advances the conversation of improving network slice mobility in next-generation mobile systems.

Conventional handover methods, such as relying on RSRP, are insufficient for meeting the varied service requirements of network slicing. In [6], researchers explore the challenges of mobility management in network slicing and highlight the complexities of handovers. They propose an innovative solution that utilizes artificial intelligence, specifically reinforcement learning, to create an intelligent handover algorithm tailored for Radio Access Network (RAN) slicing environments. The Multi-Agent Learning based Smart Handover Scheme (LESS) aims to reduce long-term handover costs while maintaining QoS. LESS is comprised of two components: LESS-DL, which selects both target base stations and network slices during handovers, and LESS-QVU, which updates Q-values through data sharing to overcome limited data availability. The suggested LESS framework has shown significant enhancements in network handover cost, handover frequency, and outage probability compared to other state-of-the-art approaches. Authors in [7] have also ventured into the dynamic deployment of network slices by introducing a prediction-assisted adaptive network slice expansion algorithm. This innovative approach comprises three core elements. Firstly, they utilize the Holt-Winters prediction algorithm to forecast network slice traffic demand, with a primary focus on minimizing disruptive changes to network topology. Secondly, their methodology incorporates a virtual network function (VNF) adaptive scaling strategy, precisely determining the appropriate number of VNFs and resources to prevent resource wastage. Lastly, they present a proactive online deployment algorithm that dynamically deploys network slices, ensuring adherence to delay requirements while considering critical factors such as resource capacity, delay constraints, network costs, and energy consumption.

The authors present a handover prediction and management technique tailored for 5G cellular networks [8]. Their approach centers on leveraging a deep learning neural network (DLNN), with a specific focus on User Equipment (UE) mobility and continuous monitoring of RF signal conditions. LSTM plays a pivotal role in their research, aiding in the tracking of UE movements. Similarly, several other works [9], [10], have delved into LSTM applications, particularly for short-term traffic predictions. Here, the deep learning LSTM model adeptly grasps long-term data dependencies and non-linear traffic patterns, aiding in making informed decisions, especially during peak traffic periods. It's worth noting that while LSTM has shown promise in handover management, its implementation has been primarily within conventional, non-sliced networks. However, there's a compelling prospect for extending LSTM-based solutions to predict network slice traffic demands, an area we will explore in our research.

III. System Model

We consider a registration area represented by R. There are a N number of tracking areas within a registration area R, represented by the set $T = \{T_1, T_2, ..., T_N\}$. Each tracking area has M number of base stations represented by $B = \{B_1, B_2, ..., B_M\}$. There are a total K number of network slices corresponding to different services within a registration area R. The available network slices are represented as $S = \{S_1, S_2, ..., S_K\}$. These network slices are randomly distributed throughout the registration area.

Each Vehicular User Equipment (VUE) is associated with a specific slice $S_K \in S$ requesting a service. The VUEs request the base station within a tracking area for its services while relying on the Cellular Vehicle-to-Everything (C-V2X) mode 4 semi-persistent scheduling (SPS) mechanism [11]. Whenever the VUE generates a request it arrives at the transport layer

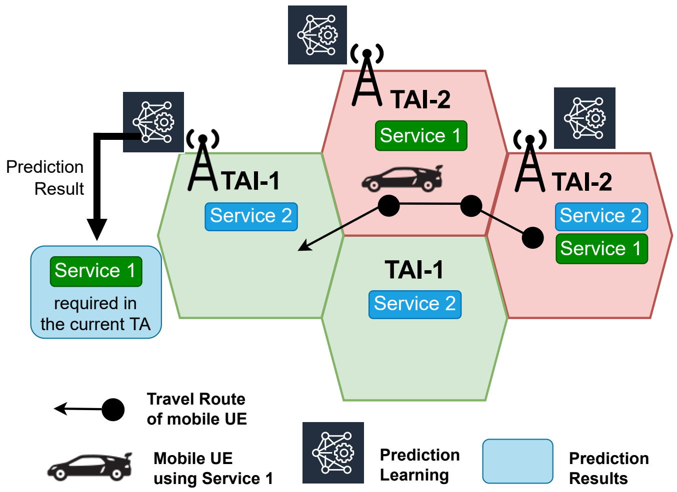


Fig. 1. Network Slice Prediction for Seamless Service.

to be sent, the VUE stores the generated packet at the buffer and first reserves the resources from the pool to transmit the generated packet. The VUE reserves the resources from the dedicated slice assigned for the specific application. Each slice is composed of a two-dimensional frequency and time grid called the resource pool. In the time domain, the sub-resource pools/slices are divided into subframes, and in the frequency domain, they are divided into subchannels. Each subframe is divided into slots of $1msec$. The total bandwidth is distributed between the slices based on the given requirements.

We have used the *deepslice* dataset for the purpose of prediction in our model [12]. A total of 3 months of data belonging to five different network slices is considered. The data includes traffic data of smart transportation, traffic safety, industry 4.0, AR/VR/gaming, and smartphone traffic. The proposed algorithm is trained using the given dataset and prediction of slice demand is carried out. The proposed system model is represented in Fig. 1.

IV. Slice Demand Prediction based Solution

A significant challenge in the realm of 5G network slicing lies in the deployment of slices that are not evenly distributed across the network. Up until Rel-17, the 3GPP operated under the assumption of uniform service support across all tracking areas within a registration area. However, this assumption does not align with practical deployment scenarios. Recognizing this, 3GPP is currently addressing the issue of non-uniform slice deployment as part of its Rel-18 efforts. The non-uniform distribution of slices can result in several unfavorable outcomes, such as service disruptions when transitioning across slice service area boundaries and inconsistent service quality throughout the registration area. Tackling the challenge of non-uniform slice deployment is essential, but the cost associated with uniformly deploying new slices across the entire registration area can be prohibitive. Therefore, it is crucial to explore cost-effective alternatives.

Efficient slice handover mechanisms emerge as promising solutions to address this challenge. Efficiently carrying out inter and intra-slice handover can play a significant role in solving the challenge of seamless service provision in the registration area due to the non-homogeneous deployment of network slices. These mechanisms can dynamically allocate network resources, negating the need for extensive uniform slice deployment. This adaptive approach can further optimize resource utilization and minimize service interruptions as users and devices move across different tracking areas within a registration area. However, to effectively implement such an approach, it becomes highly advantageous to anticipate slice demand. Such proactive resource management can allow for the seamless provision of services to users within a registration area. In light of this, we present a solution based on Recurrent Neural Networks (RNNs) to forecast the demand of existing network slices. Our proposed approach harnesses LSTM networks to make predictions regarding slice demands. These predictive insights into network slice data enable us to dynamically allocate resources among the available slices. This resource adjustment accommodates incoming traffic from neighboring tracking areas, effectively addressing the mobility management challenge within the context of network slicing. For instance, when a new user seeks access to a specific service not currently accessible in their target tracking area (the tracking area the user is moving towards), the allocated resources in that area can be preemptively reconfigured. This proactive resource reallocation is guided by the predictions generated by our LSTM-based model, allowing us to provide the requested service to the incoming user. Our proposed scheme not only assists in solving the challenges of mobility management and non-homogeneous deployment of network slices within a registration area but also injects a layer of intelligence into network operations, aligning them with the 5G-Advanced criteria set forth by 3GPP.

TABLE I
LSTM Neural Network Parameters

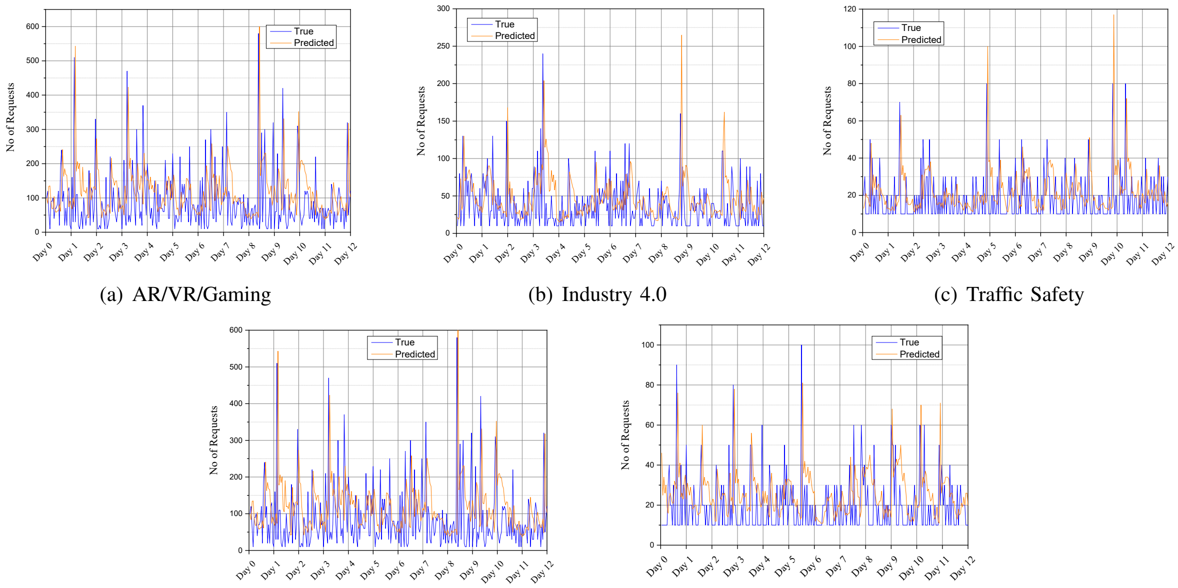
Parameter	Value
Hidden Layers	2
No. of Neurons in each layer	64
Learning Rate	0.001
Batch size	256
Dropout	20%
Activation Function	Sigmoid

V. Evaluation

In this section, we describe our simulation setup employed to carry out network slice demand prediction using LSTM. Furthermore, we present the results of the prediction for available network slices.

A. Simulation Setup

For the purpose of evaluation, we have used *'deepslice'* dataset [12]. We have performed preprocessing on the dataset to use it for prediction of slice demands. The dataset includes



(a) AR/VR/Gaming (b) Industry 4.0 (c) Traffic Safety

(d) Smartphone (e) Smart Transportation

Fig. 2. Prediction of received number of requests for different network slices.

smartphone, traffic safety, industry 4.0, AR/VR/Gaming, and smart-transportation slices. We have considered three months of data which includes incoming traffic of VUEs within one registration area where VUEs request different services from the available network slices. The simulation setup follows our system model presented in section III. The dataset is divided into 85/15 split for training and testing of our LSTM model. The model is trained for 300 epochs. The parameters of our LSTM model are presented in Table I.

B. Simulation Results

This section discusses the prediction results obtained using our LSTM model. Firstly, the individual prediction results of the available network slices' traffic demand are discussed. Then, the average slice traffic demand for each service is compared. Finally, we analyze the training performance of the LSTM neural network.

Fig. 2 presents the predictive outcomes pertaining to the available network slices. Specifically, the forecast spans a 12-day period, encompassing the user-generated request statistics for each network slice service. The graphical depictions within Fig. 2(a) through 2(e) are dedicated to the services of AR/VR/Gaming, Industry 4.0, Traffic Safety, Smartphone, and Smart Transportation, respectively. Within each subfigure, a comparative visualization unfolds, showcasing both the actual and predicted request quantities corresponding to the respective service. Evidently, the LSTM model demonstrates a remarkable capacity for precise prediction across all services. These meticulously forecasted datasets, tailored to each service, offer invaluable insights into slice traffic demand. Subsequently, these insights empower proactive resource allocation or reallocation among network slices, facilitating the seamless accommodation of incoming users originating from neighboring tracking areas.

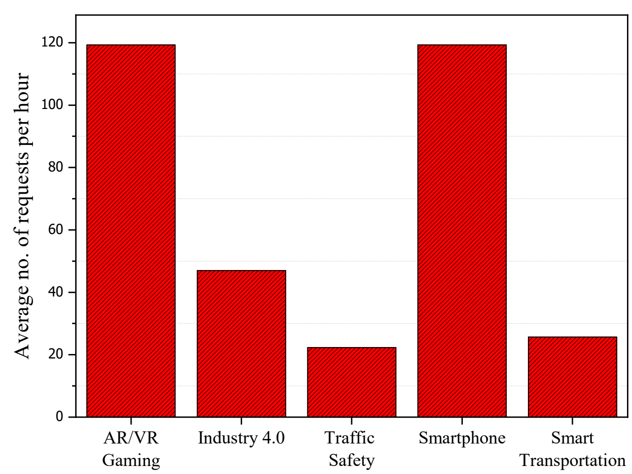


Fig. 3. Average slice traffic demand per hour.

A total of five network slices were considered for predicting the number of requests received every hour for a 12-day temporal span. Using the predicted data, we infer the average slice demand per hour for each service considering the fluctuations in the volume of requests received every hour. The plot in Fig. 3 exhibits the highest demand for AR/VR/Gaming and Smartphone slice whereas the lowest demand for Traffic Safety slice considering the number of requests received for respective slices. This information can be used to proactively (re)configure the resources among slices in order to provide seamless service to the incoming traffic in the current tracking area. Consequently, we can ensure seamless service provision for the users within a registration area and efficiently tackle the challenges of mobility management and non-uniform slice deployment throughout the registration area.

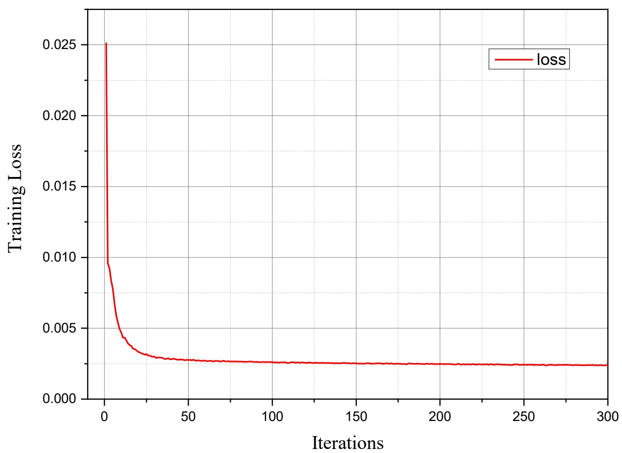


Fig. 4. Training loss of LSTM model.

Finally, in Fig. 4, we illustrate the training loss curve of our LSTM neural network model. This graphical representation serves as an insightful analysis of the model's convergence, leveraging the dataset of network slice traffic for various available services. The discerning observation from this plot is the progressive reduction in loss values as the number of epochs increases. Remarkably, these loss values approach near-zero levels within the initial 100 epochs. This swift convergence can be attributed to the dataset's relatively modest size and simplicity, which enables the model to swiftly reach its optimal state. Consequently, the model exhibits quick convergence accompanied by stable training performance.

VI. Conclusions and Future Work

In this manuscript, we have presented a solution addressing the complexities of mobility management and non-uniform network slice deployment. Our approach leverages predictive learning, employing Long Short-Term Memory (LSTM) to forecast request volumes for various network slices. These predictions enable us to deduce the demand for each slice, facilitating proactive resource allocation to accommodate future demand shifts. This proactive configuration ensures uninterrupted service delivery to users transitioning from neighboring tracking areas, thus guaranteeing seamless connectivity across the entire registration area. Our proposed method can effectively minimize service disruptions for mobile users while simultaneously tackling the challenge of non-uniform slice deployment.

Moving forward, our future endeavors will focus on developing a robust handover mechanism to address both inter and intra-slice challenges by utilizing the prediction results. Additionally, we plan to implement an efficient resource management scheme tailored to diverse tracking areas within registration areas.

VII. Acknowledgement

This research has been supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government(MSIT), (NRF-2022R1A2C1003620). This work was also supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korean Government(Ministry of Science and ICT). (RS-2023-00304695).

References

[1] R. A. Addad, T. Taleb, H. Flinck, M. Bagaa, and D. Dutra, "Network slice mobility in next generation mobile systems: Challenges and potential solutions," *IEEE Network*, vol. 34, no. 1, pp. 84–93, 2020.

[2] M. M. Sajjad, C. J. Bernardos, D. Jayalath, and Y.-C. Tian, "Inter-slice mobility management in 5g: motivations, standard principles, challenges, and research directions," *IEEE Communications Standards Magazine*, vol. 6, no. 1, pp. 93–100, 2022.

[3] "Nokia; 5g-advanced: Expanding 5g for the connected world, author=NOKIA, year=2022."

[4] M. A. Tairq, M. M. Saad, M. T. R. Khan, J. Seo, and D. Kim, "Drl-based resource management in network slicing for vehicular applications," *ICT Express*, 2023.

[5] N. A. Mohammedali, T. Kanakis, M. O. Agyeman, and A. Al-Sherbaz, "A survey of mobility management as a service in real-time inter/intra slice control," *IEEE Access*, vol. 9, pp. 62533–62552, 2021.

[6] Y. Sun, W. Jiang, G. Feng, P. V. Klaine, L. Zhang, M. A. Imran, and Y.-C. Liang, "Efficient handover mechanism for radio access network slicing by exploiting distributed learning," *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2620–2633, 2020.

[7] J. Zhou, W. Zhao, and S. Chen, "Dynamic network slice scaling assisted by prediction in 5g network," *IEEE Access*, vol. 8, pp. 133700–133712, 2020.

[8] R. A. Paropkari, A. Thantharate, and C. Beard, "Deep-mobility: A deep learning approach for an efficient and reliable 5g handover," in *2022 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, pp. 244–250, IEEE, 2022.

[9] P. Poonia and V. Jain, "Short-term traffic flow prediction: using lstm," in *2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3)*, pp. 1–4, IEEE, 2020.

[10] A. Alamoudi, M. Liu, A. Payani, F. Fekri, and D. Li, "Predicting mobile users traffic and access-time behavior using recurrent neural networks," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6, 2021.

[11] M. M. Saad, M. T. R. Khan, S. H. A. Shah, and D. Kim, "Advancements in vehicular communication technologies: C-v2x and nr-v2x comparison," *IEEE Communications Magazine*, vol. 59, no. 8, pp. 107–113, 2021.

[12] A. Thantharate, R. Paropkari, V. Walunj, and C. Beard, "Deepslice: A deep learning approach towards an efficient and reliable network slicing in 5g networks," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, pp. 0762–0767, IEEE, 2019.