Dynamic Resource Block Allocation in Network Slicing

Nidhi *‡, Bahram Khan†§, Albena Mihovska *¶, Ramjee Prasad *‖,
Vladimir Poulkov ‡††, Fernando J. Velez †**
*AU – Department of Business Development and Technology, Aarhus University, Herning, Denmark
†Instituto de Telecomunicações and Universidade da Beira Interior,
Faculdade de Engenharia, Departamento de Engenharia Electromecânica Covilhã, Portugal
‡TU – Technical University Sofia, Bulgaria
Emails: ‡nidhi@btech.au.dk, §bahram.khan@lx.it.pt, ¶amihovska@btech.au.dk, ‖ramjee@btech.au.dk,
††vkp@tu-sofia.bg, **fjv@ubi.pt

***Abstract*—Network slicing is crucial in 5G and its evolution concerning user-centric services. By allocating independent resources, like link bandwidth, computing/processing capabilities and spectrum, to address users' requests, slicing serves the end-to-end verticals or services. gNodeB (gNB) allocates the bandwidth resources to transmit/receive data to User Equipments (UEs). Resources Blocks (RBs) are the smallest resource entities assigned to a single user. In 5G New Radio (NR), the time-domain resource allocation defines the allocated symbols (OFDM symbols), while the frequency-domain allocation illustrates the RBs (sub-carriers) allocation to the UE. RB comprehends 12 sub-carriers in the frequency domain with a flexible RB bandwidth, unlike LTE-A. It is critical to provide enhanced services to different users. There have been several works on challenges to enable a multi-tenant and service RAN while providing isolation to the slices. This work proposes a detailed approach for creating slices based on the demanded services, resource virtualization and isolation. The focus is on resource sharing algorithms at the Slice Orchestrator (SO) level. These virtual network slices support a wide range of services and applications categorized into the Enhanced Mobile Broadband, Ultra-Reliable and Low-Latency Communications and Massive Machine-Type Communications megatrends. The paper also provides an overview of standardization activities and evolving requirements to support use cases and services like Holographic Telepresence, Automotives, among other.**

***Index Terms*—Network Slicing, Communication Service Providers, RB Allocation, Slice Isolation, Holographic Communication, Industry 4.0.**

I. Introduction

The roll-out of fifth-generation (5G) wireless networks has embarked on the concept of Network Slicing (NS) as one of its fundamental technologies. NS facilitates a segmented layer of networks as slices or network slices in addition to the base network architecture [1]. As an integral part of the virtual network, these network slices offer full end-to-end connectivity for user-specific services [2]. The slices forming isolated virtual network layers exhibit all the required functionalities of the shared physical network. It enables the Communication Service Providers (CSPs) to address various inventive business models, use cases, and tailor-made user-specific solutions with guaranteed performance over a prevailing infrastructure. NS diminishes the requirement for new physical networks for dedicated services. The network orchestration helps CSPs to automate the communication on and across the network providing tailored services with guaranteed Quality of Service (QoS) for various usage scenarios, in association with Service Level Agreements (SLAs) [3].

NS Framework was first proposed by the Third Generation Partnership Project (3GPP) in Release 15 [4]. It has gradually evolved, with the technical details and enhancements in the following releases. By employing NS, CSPs can create multiple virtual slices to acknowledge enormous data traffic increases and specific user requirements. Each slice in isolation, i.e., without interfering with the coexisting slices, hosts individual network functions and application services [5]. Various Standard Development Organizations (SDOs) have backed NS to support multi-vendor services [6]. The Next-Generation Mobile Network (NGMN) [1] laid out the detailed principle of creating and managing multiple independent logical mobile networks over shared physical infrastructure.

This work focuses on resource-sharing algorithms at the Slice Orchestrator (SO) level and computing slice's radio resource requirements and usage. The resources are periodically adjusted based on the current Channel Quality Indicator (CQI). The scheduler needs to handle the demand and service quality. As it is essential to efficiently allocate radio resources in dynamic environments, allocating resources to slices distinguishes services to meet user QoS. The proposed method tracks adaptive behaviors of communication services based on the number of active users, data buffer status, and channel condition. Additionally, the paper also provides different use cases enabling future networks and an overview of the standardization activities towards NS.

The remaining paper is organized as follows. Section II discusses the state-of-the-art concerning resource management, NS standardization activities, and requirements. Section III

This work was supported by FCT/MCTES through national funds and when applicable cofounded EU funds under the project UIDB/50008/2020, ORCIP (22141-01/SAICT/2016), COST CA 20120 INTERACT, SNF Scientific Exchange-AISpectrum (project 205842) and TeamUp5G. TeamUp5G has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie ETN TeamUp5G, grant agreement No. 813391.

summarizes our methodology and assumptions, along with the discussion of the results. Finally, Section IV presents the main conclusions of this work and discusses topics for further research.

II. State-of-the-Art

The evolving services are categorized broadly into the following four classes: (i) enhanced Mobile Broadband (eMBB), (ii) Ultra-Reliable Low-Latency Communications (URLLC), (iii) massive Machine-Type Communications (mMTC), and (iv) Vehicle to Everything (V2X). These services set the Key Parameter Indicators (KPIs) to evaluate the user requirements enabled through network slicing. In 4G networks, NS was limited to service isolation [7]. However, with 5G and beyond networks, NS can facilitate CSPs to provide guaranteed QoS services via virtual network slices, also called "5G Slice". The resources are allocated on-demand with a valid session using existing virtualization techniques. Thus, it adds additional scalability and flexibility to conventional networks. By facilitating resource sharing and automation among independent 5G slices, virtualization and Orchestration technologies are the key enablers of NS [8], [9]. NGMN defined a 3-layer NS framework [1] enabling a flexible and scalable End-to-End (E2E) architecture, consisting of the Radio Access Network (RAN) and core networks [10]. 5G Infrastructure Public-Private Partnership (5G-PPP) proposed an exhaustive 5-layer NS Framework in addition to the NGMN definition, depending on different use cases [6], [11]. To manage the slice sessions and resource management, ETSI defined the Network Function Virtualization (NFV) Management and Network Orchestration (MANO) architecture [9] to facilitate NF management and orchestration, its virtualization and resource allocation through the three following functional blocks: i) the NFV orchestrator, ii) the VNF manager, and iii) virtualized infrastructure manager [12]. Figure 1 illustrates the various layers that conceptualize the NS technology.

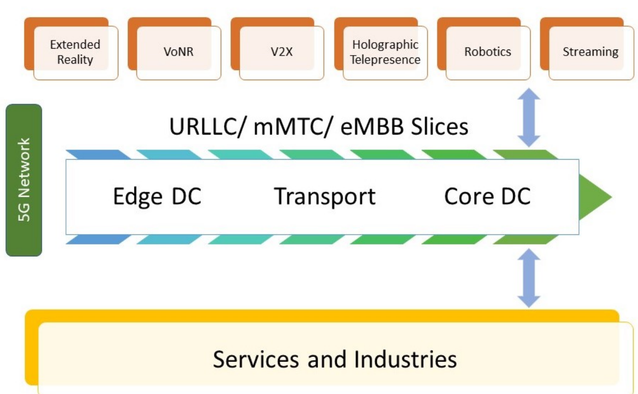

Fig. 1. Layered approach and concept of Network Slicing

A. Resource Block Allocation in Network Slicing

In wireless networks, Dynamic Resource Allocation (DSA) of data packets is critical to support different dynamically active services like real-time (RT), non-real-time (NRT), and control signaling. The gNodeB (gNB) allocates the bandwidth resources to the User Equipment (UE) to facilitate data transmission and reception in both downlink and uplink. The Resource Block (RB) is the smallest resource entity assigned to a single user. The time-domain resource allocation in 5G New Radio (5G NR) defines the allocated symbols (OFDM symbols) from different sub-carriers. In contrast, the frequency domain allocation illustrates the Resource Block (RB) (sub-carriers) allocation to the UE. An RB comprehends 12 sub-carriers in the frequency domain with a flexible RB bandwidth, unlike LTE-A. RB bandwidth depends on sub-carrier spacing. NR provides a higher bandwidth efficiency (up to 99%) than the LTE (90%) [13] and operates at a channel bandwidth of 100 MHz, in the sub-6 GHz bands, and 400 MHz, in the mmWave range, without any reserved Direct Current (DC, the sub-carrier whose frequency is equal to the RF center frequency of the transmitting station) sub-carrier for uplink and downlink. The UEs use a DC subcarrier to identify the OFDM frequency band's center and do not contain any information. The maximum and minimum RBs are defined by means of the 5G New Radio numerology. Hence, the channel bandwidth can be calculated by knowing the given bandwidth of the RB.

Finally, it is worthwhile to mention that the 3GPP specifications define two types of resource allocation in the frequency domain:

- Resource Allocation Type #0 (RAT#0);
- Resource Allocation Type #1 (RAT#1).

Network Slicing scenario deployed with slices created over the same infrastructure demands sharing the same network resources. The Slice Manager (SM) is responsible for allocating resources to individual slices while coordinating with the infrastructure providers. The Virtual Network Operator (VNO) services the slices, commonly known as the slice tenants. SLA is created between SMs and VNOs to regulate the required resources [14]. Authors in [15] have presented an efficient approach for statistical resource distribution among the network slices with strong SLAs. This approach provided a higher trade-off between the resource distribution and system complexity and thus, opened new research questions on the data and cost continuum. In [14], the authors addressed the slicing of RAN resources in multi-tenant scenarios. The resource allocation approach focused on the optimized fairness index, utility gains, and capacity savings. Following this, authors in [16] explored the Deep Reinforcement Learning (DRL) approach to allocate resources in dynamic multi-tenant systems. In [17], authors have proposed an adaptable and flexible 5G network architecture to support cross-domain E2E slicing with well-defined inter-slice control and management functions.

B. Standardization Activities on Network Slicing

Standards administrate the development of products and technologies to ensure requirements, interoperability, and quality [18]. Organizations like Global System for Mobile Communications Association (GSMA) and NGMN have gradually contributed to the high-level system requirements and architecture for Network Slicing. They also regulate the fundamentals for creating slices in an E2E 5G NS framework [19]. GSMA [20] highlighted the need for collaboration in the standardization process from the giants of different verticals, namely academia, industries, CSPs, etc.

3GPP is actively involved in multiple initiatives to support 5G network slicing like SA1 (requirements and use cases), SA2 (NS Architecture) [19], SA3 (Security), SA5 (Slice Management) [21], etc. Similarly, the Internet Engineering Task Force (IETF) contributes to the requirements and applications. The European Telecommunications Standards Institute (ETSI) is working towards NS services, configuration, delivery, assurance of deployment, etc. [9]. The 13th study group (SG13) of International Telecommunication Union - Telecommunication (ITU-T) focuses mainly on the orchestration, network management, and horizontal slicing [22]. Furthermore, the focus group FG-ML 5G defines Machine Learning developments and scopes to the requirements and services [23]. Authors from [6] have mentioned different relevant groups working on NS standardization activities.

C. Requirements for Future Networks

The use cases, service requirements, business models and application areas constantly evolve to meet the diverse demands. This sub-section lists some characteristic use cases and services for 5G and beyond networks [24].

- The underlying requirements are as follows:
  - Holographic Communication - Holographic Telepresence (HT) or communication is the next frontier to provide an immersive experience of distant communication with or without using the Head Mount Devices (HMDs) [25]. It focuses on amalgamating sensory information like touch, smell, and taste into the audio and visual transmission and reception. To facilitate HT, ultra-low latency of 1ms [26] and ultra-high data rate (Tbps) to support 30fps [27] are required, with high computing capabilities.
  - Industrial Automation - The upcoming generation of mobile communication foreseen an industrial revolution and dominance of automated services. The everything-to-everything connection will rule the networking paradigm with an enormous amount of devices. The services demand high QoS and Quality of Experience (QoE). The requirements to support Industry 4.0 and future 5.0 will be Ultra-low latency of 0.1 ms, ultra-high reliability, and ultra-low delay jitter [26], [28].

Figure 2 illustrates the classes of use cases for upcoming communication networks [29].

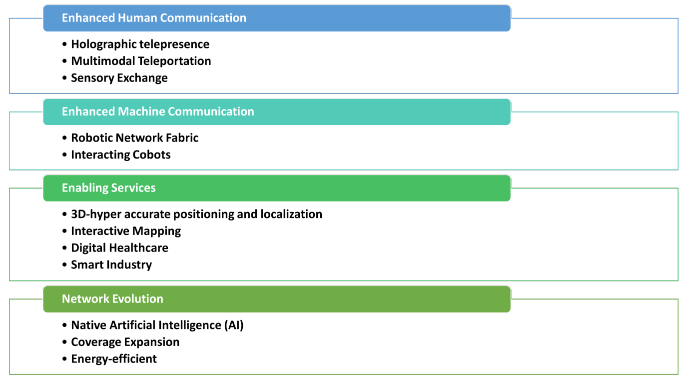

Fig. 2. High-level Use Cases Categorization [29]

III. Methodology and Results

Network slicing enabled the Infrastructure/Slice Providers (SP) to offer resources to the customers as a service for a given cost to maximize the resource usage by accepting slice requests. The customer requests a network slice from SP to get customized services. There is a need for a mechanism/ scheduling scheme in which SPs can entertain the requests, as the SPs are subject to limited resources. Authors from [15] and [30] introduced a two-level scheduler to share the Physical Radio Blocks (pRB) among slices by abstracting pRBs and using two scheduler levels. Two-level schedulers operation is as follows:

- The first level is slice-specific, allowing each slice to use its internal scheduler and schedule each UE with Virtual Resource Blocks (vRB).
- The second level considers the slice-specific (virtual) resource assignment and maps it to actual pRBs. It controls the number of NpRBs (number of pRBs) assigned to each slice and indicates the maximum NpRB to dedicate to each slice after executing an intra-slice physical resource sharing algorithm.

The aim is to compute the radio resources required in each slice. The resources are periodically adjusted based on the current Channel Quality Indicator (CQI) estimates from the users of the different slices. Assumptions are as follows:

- 5G network which including a SO, to initiate and configure slice resources based on the use case types (eMBB, mMTC or URLLC) and a set of eNBs deployed covering an area.
- The SO communicates with the eNBs using a protocol that allows remote interaction and management.
- The eNB management process consists of
  - RAN information (CQI);
  - eNB configuration.
- A set of UEs is served by/associated with a network slice, spanning a set of eNBs (i.e., different physical locations).
- There are three types of Slices: eMBB and URLLC Slices
- The SO receives the request to instantiate a slice. The Slice request includes
  - Slice type;

– Duration;
– Requirements like data rate, application, or latency;
– List of associated UEs.

We simulated on MATLAB the two-level scheduler for eMBB and URLLC slices for a varying number of users in each slice, keeping the other constant to observe CQI variations. The goal is to improve network performance and introduce flexibility and optimization of the network resources by accurately and dynamically provisioning the activated network slices with the appropriate resources to meet their diverse requirements. The aim is to have a flexibility in RAN resource allocation concerning slicing.

A. Slice Definition and Requirements

- eMBB Slice Requirements - High Data Rate

$$N_{pRB_{max}}(i) * d_{pRB} = N_{users}(i) * d_{App/user}$$

- URLLC Slice Requirements: Ultra-Low Latency

$$\mu = \frac{N_{pRB} * d_{pRB}}{\text{Average packet size}}$$

where,
$N_{pRB_{max}}(i)$: Required pRBs for each eNB;
$d_{App/user}$: Required Data rate per slice.;
N_{users} : the number of active users;
d_{pRB} : maximum data rate provided by one pRB;
pRB : Physical Radio Blocks.
Ideal channel conditions correspond to the maximum CQI = 15.

B. Simulation Results and Discussion

The performance has been evaluated in Matlab as an extension to the referenced work in [15]. We modified the SO and considered eMBB and URLLC slice. We defined each slice with the required data rate, number of users and latency in URLLC slice. The simulations were carried out at varying CQI level, i.e. ,medium (7) to high (13). The eMBB slice users were fixed to 5 and varied users (up to 20 for medium CQI and up to 30 for high CQI values) in URLLC slice.

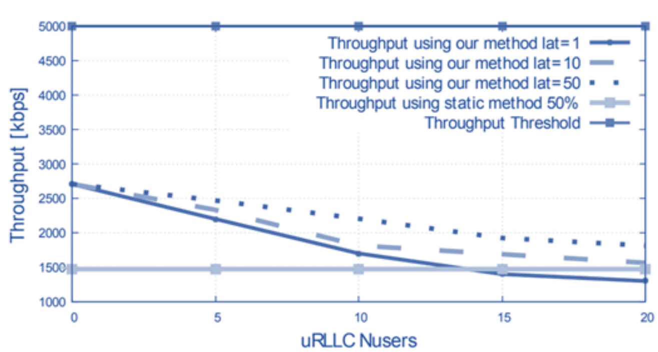

Fig. 3. Throughput variations as a function of the number of users (with Medium CQI value).

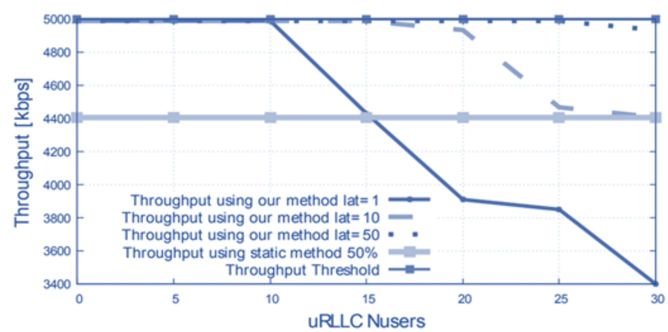

Fig. 4. Throughput variations as a function of the number of users (with High CQI value).

Figures 3 and 4 present the throughput for the URLLC slice at varying CQI values. Beyond the threshold, slice performance degrades (more in case of high CQI) but it guarantees the required bandwidth until 25 users.

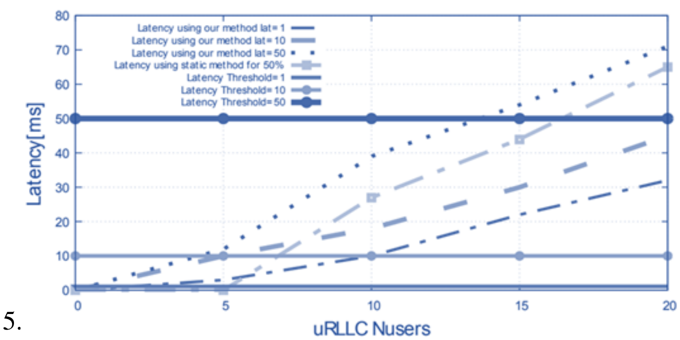

Fig. 5. Latency variations as a function of the number of users (with Medium CQI value).

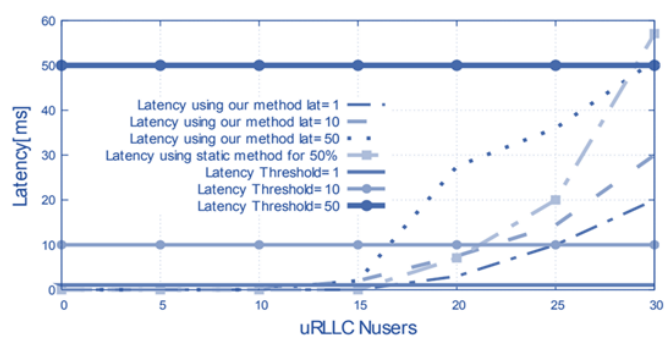

Fig. 6. Latency variations as a function of the number of users (with High CQI value).

Figures 5 and 6 illustrate the experienced latency in case of the uRLLC users for varying CQI values from 7 to 13. We considered different maximum values for latency, i.e., 1 ms, 10 ms and 50 ms. We observed that the max. latency value was near about maintained for both the value of CQI. Good CQI allows higher NpRB compared with the medium CQI by allowing more users. We also observed that fixed number of pRBs cannot guarantee the very low latency requirement.

The results in Figures 7 and 8 show the estimated NpRB and used by the eNBs for the URLLC slices with varied CQI values. In 7 and 8, we observed that the estimated NpRB is similar to the one communicated to the eNB until reaching the identified thresholds as in Figures 4 and 5. The communicated NpRB to eNB is lower than the estimated value on increasing the threshold values.

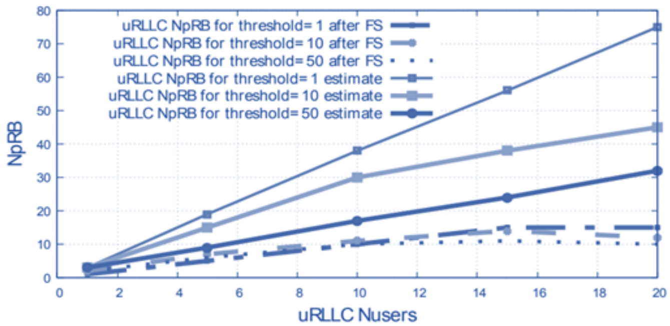

Fig. 7. NpRB of the URLLC slice as a function of the number of users.

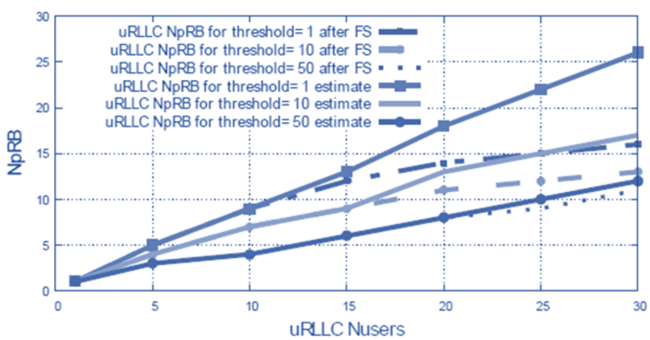

Fig. 8. NpRB of the URLLC slice as a function of the number of users.

The results in Figures 9 and 10 show the estimated NpRB used by the eNBs for eMBB slices with varied CQI values. We observed that the estimated NpRB could not be satisfied in case of medium channel quality and required NpRB is higher. Lower dpRB is expected at higher CQI values.

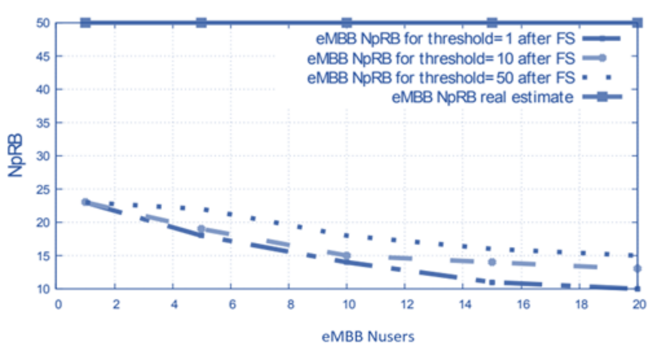

Fig. 9. NpRB of the eMBB slice as a function of the number of users.

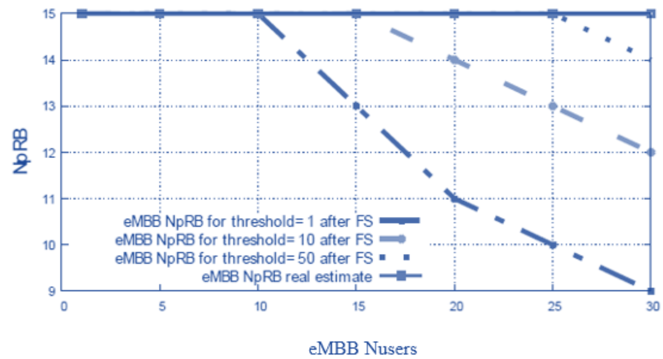

Fig. 10. NpRB of the eMBB slice as a function of the number of users.

The results show that our proposed algorithm to estimate the required NpRB for eMBB and URLLC slices is accurate and permits sharing of the RAN resources among slices. Hence, the practical feasibility of our proposed solution is verified.

IV. Conclusions and Future Work

This paper proposes slice creation and allocation of resource blocks while isolating the slices for eMBB and URLLC by using the two-level scheduling introduced in the referenced work. We have introduced algorithms to estimate the required RAN resources for the eMBB and URLLC slices while evaluating the performance under varying CQI values at the SO level. Heterogeneous Networks (HetNets) optimization is an open research area concerning Network Slicing. Also, 3GPP's functional splits [20] have huge potentials to be implemented with slicing to manage the network function virtualization and softwarization of RAN resources. Besides, there is a need to develop novel meta-learning models for ML-enabled network slicing, an open research area.

The main foreseen challenge in 5G New Radio dynamic resource allocation is the associated overhead when we extract the information from the base station (UE provides CQI to the base stations) to the SO. Thus, to eliminate/ minimize the communication overhead, we will simultaneously propose the following steps:

- A machine learning approach to infer the stability of UE channel conditions;
- Propose a predictive scheme to efficiently reduce the dependency on the network's configuration to address the various service and demands;
- Admission Control Policy/ Decision based on Q-Learning and Regret Matching for the SP to manage the slice requests (we will then validate the mechanism concerning the SP serving the network requests).

References

[1] N. Alliance, "5G White Paper (Final Deliverable)," 2015.

[2] H. Zhang, N. Liu, X. Chu, K. Long, A.-H. Aghvami, and V. C. Leung, "Network slicing based 5G and future mobile networks: mobility, resource management, and challenges," *IEEE communications magazine*, vol. 55, no. 8, pp. 138–145, 2017.

[3] "Network slicing explained," Nov 2020. [Online]. Available: https://www.nokia.com/about-us/newsroom/articles/network-slicing-explained/
[4] K. Flynn. (2019) A global partnership. [Online]. Available: https://www.3gpp.org/release-15
[5] A. Nakao, P. Du, Y. Kiriha, F. Granelli, A. A. Gebremariam, T. Taleb, and M. Bagaa, "End-to-end network slicing for 5g mobile networks," *Journal of Information Processing*, vol. 25, pp. 153–163, 2017.
[6] B. Khan, Nidhi, A. Mihovska, R. Prasad, and F. J. Velez, "Overview of network slicing: Business and standards perspective for beyond 5g networks," in *2021 IEEE Conference on Standards for Communications and Networking (CSCN)*, 2021, pp. 142–147.
[7] Samsung, "Samsung Technical White Paper: Network Slicing," 2020.
[8] N. Alliance, "Description of network slicing concept," *NGMN 5G P*, vol. 1, no. 1, 2016.
[9] N. ETSI, "Network functions virtualisation (NFV); terminology for main concepts in NFV," *Group Specification, Dec*, 2014.
[10] X. Foukas, G. Patounas, A. Elmokashfi, and M. K. Marina, "Network Slicing in 5G: Survey and Challenges," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 94–100, 2017.
[11] S. Redana, Ö. Bulakci, A. Zafeiropoulos, A. Gavras, A. Tzanakaki, A. Albanese, A. Kousaridas, A. Weit, B. Sayadi, B. T. Jou *et al.*, "5G PPP architecture working group: View on 5G architecture," 2019.
[12] A. Devlic, A. Hamidian, D. Liang, M. Eriksson, A. Consoli, and J. Lundstedt, "NESMO: Network slicing management and orchestration framework," in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2017, pp. 1202–1208.
[13] X. You, C. Zhang, X. Tan, S. Jin, and H. Wu, "Ai for 5g: research directions and paradigms," *Science China Information Sciences*, vol. 62, no. 2, pp. 1–13, 2019.
[14] P. Caballero, A. Banchs, G. de Veciana, and X. Costa-Pérez, "Multi-tenant radio access network slicing: Statistical multiplexing of spatial loads," *IEEE/ACM Transactions on Networking*, vol. 25, no. 5, pp. 3044–3058, 2017.
[15] A. Ksentini, P. A. Frangoudis, P. Amogh, and N. Nikaein, "Providing low latency guarantees for slicing-ready 5g systems via two-level mac scheduling," *IEEE Network*, vol. 32, no. 6, pp. 116–123, 2018.
[16] F. Mason, G. Nencioni, and A. Zanella, "A multi-agent reinforcement learning architecture for network slicing orchestration," in *2021 19th Mediterranean Communication and Computer Networking Conference (MedComNet)*. IEEE, 2021, pp. 1–8.
[17] M. Shariat, Ö. Bulakci, A. De Domenico, C. Mannweiler, M. Gramaglia, Q. Wei, A. Gopalasingham, E. Pateromichelakis, F. Moggio, D. Tsolkas *et al.*, "A flexible network architecture for 5g systems," *Wireless Communications and Mobile Computing*, vol. 2019, 2019.
[18] N. Abdelkafi, R. Bolla, C. J. Lanting, A. Rodriguez-Ascaso, M. Thuns, and M. Wetterwald, "Understanding ICT standardization: Principles and practice," 2019.
[19] P. Rost, C. Mannweiler, D. S. Michalopoulos, C. Sartori, V. Sciancalepore, N. Sastry, O. Holland, S. Tayade, B. Han, D. Bega *et al.*, "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 72–79, 2017.
[20] GSMA. (2021) E2E Network Slicing Architecture Version 1.0. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads//NG.127-v1.0-2.pdf
[21] X. De Foy and A. Rahman, "Network slicing-3GPP use case," *Working Draft, IETF Secretariat, Internet-Draft draft-defoy-netslices-3gpp-network-slicing-02*, 2017.
[22] ITU-T. (2019) Progress of 5G studies in ITU-T: overview of SG13 standardization activities . [Online]. Available: https://www.itu.int/en/ITU-T/Workshops-and-Seminars/20180604/ Documents/Session1.pdf
[23] "Focus Group on Machine Learning for Future Networks including 5G." [Online]. Available: https://www.itu.int/en/ITU-T/focusgroups/ml5g/Pages/default.aspx
[24] Y. Zhou, L. Liu, L. Wang, N. Hui, X. Cui, J. Wu, Y. Peng, Y. Qi, and C. Xing, "Service-aware 6g: An intelligent and open network based on the convergence of communication, computing and caching," *Digital Communications and Networks*, vol. 6, no. 3, pp. 253–260, 2020.
[25] A. Vlahov, V. Poulkov, and A. Mihovska, "Analysis of open ran performance indicators related to holographic telepresence communications," in *2021 24th International Symposium on Wireless Personal Multimedia Communications (WPMC)*. IEEE, 2021, pp. 1–5.
[26] C. Han, Y. Wu, Z. Chen *et al.*, "Network 2030 a blueprint of technology applications and market drivers towards the year 2030 and beyond," 2018.
[27] R. Li *et al.*, "Towards a new internet for the year 2030 and beyond," in *Proc. 3rd Annu. ITU IMT-2020/5G Workshop Demo Day*, 2018, pp. 1–21.
[28] G. Berardinelli, N. H. Mahmood, I. Rodriguez, and P. Mogensen, "Beyond 5g wireless irt for industry 4.0: Design principles and spectrum aspects," in *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2018, pp. 1–6.
[29] N. Alliance, "6G Use Cases and Analysis," 2022.
[30] A. Ksentini and N. Nikaein, "Toward enforcing network slicing on ran: Flexibility and resources abstraction," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 102–108, 2017.