AI-empowered IBN for Intelligent Automation of Beyond 5G Services

Khizar Abbas, Ali Nauman, Talha Ahmed Khan, Muhammad Afaq, Jae-Hyung Yoo,
James Won-Ki Hong, *Senior Member, IEEE,* and Wang-Cheol Song*

***Abstract*—Accommodating novel and diverse services with different quality of service requirements (QoS), including Web 3.0, metaverse, and augmented and virtual reality, is very challenging for Beyond 5G (B5G) networks. These services require high-speed, low latency, and seamless connectivity. Conversely, network slicing in B5G networks is the best candidate to support these services by providing dedicated resources. However, automatically orchestrating and managing network slicing for these services is challenging. Therefore, this article presents a closed-loop Intent-Based Networking (IBN) architecture for network slice design and orchestration over multidomain infrastructure. IBN automatically performs commissioning, activating, monitoring, and decommissioning end-to-end (e2e) slices. Moreover, we have integrated an Artificial Intelligence (AI) empowered network data analytics mechanism with IBN to enable network intelligence for proactive management and control. Finally, we performed several experiments by creating e2e slices through the IBN mechanism, which shows satisfactory performance in automated service provisioning, resource allocation, and resource management.**

***Index Terms*—5G/6G, AI, Network Slicing, Edge-Cloud MANO, IBN, NWDAF, Web 3.0**

I. INTRODUCTION

The main focus of the next-generation internet (web 3.0) is user-centricity and decentralization. [1]. These decentralized and innovative services are augmented reality (AR), virtual reality (VR), digital twins, smart industries, Internet of Things (IoT), smart education, and smart healthcare. The present mobile networks do not fulfill the strict service requirements of these novel decentralized services, such as ultra-low latency, high speed, and seamless connectivity. The upcoming 5G/6G network is expected to support these decentralized services. These services are majorly divided into massive machine-type communication (mMTC), ultra-reliable low latency (URLLC), and enhanced mobile broadband (eMBB). Network slicing is the main pillar of 5G architecture that enables mobile network operators (MNOs) to accommodate each service by partitioning the same physical infrastructure into various logically isolated networks. Software-defined networking (SDN) and network function virtualization (NFV) enables network slicing by allowing network programmability and virtualization [2]. So, network slicing empowers MNOs to fulfill diverse service requirements and guarantee QoS.

Automating network slicing for services in a multi-domain network infrastructure is critical because each domain needs specific policy configurations to activate resources. Moreover, the third-generation partnership project (3GPP) categorizes slice life-cycle management into commissioning, activation, monitoring, and decommissioning. The slice template is prepared in the slice commissioning step by converting the received slice requirements into network policies. The resources are deployed in the slice activation step according to the prepared slice template. Activated slice resources are continuously monitored in the monitoring phase to maintain QoS assurance. The activated slice resources are also destroyed and released during the slice decommissioning part [3]. However, the automatic orchestration and management of these novel services with differentiated quality of service (QoS) requirements is a big challenge for current and future mobile network operators (MNOs). However, it needs an automation platform that automatically manages the network slice lifecycle by generating error-free complex network configurations and deploying the resources according to the strict QoS requirements.

Besides, several industrial solutions for network slicing automation and management have used Artificial Intelligence (AI) technologies for proactive management and assurance of services. So, in 5G service-based architecture (SBA) architecture, 3GPP has presented an AI-enabled network data analytics function (NWDAF) for providing intelligence to the network. So, NWDAF provides proactive resource management and assurance using AI methods. However, a novel solution is required to enable the automatic provisioning and management of novel services over multi-tier infrastructure. It can also provide network intelligence using various innovative AI algorithms.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. NRF-2022R1A4A1032361). This work was partly supported by the Institute of Information Communications Technology Planning Evaluation (IITP) grant funded by the Korean Government (MSIT) (2018-0-00749, Development of Virtual Network Management Technology based on Artificial Intelligence) and the ITRC (Information Technology Research Center) support program (IITP-2022-2017-0-01633).

Khizar Abbas is with the Department of Computer Science, Hanyang University, Seoul, South Korea (engr.khizarabbas14@gmail.com).

Khizar Abbas, Jae-Hyoung Yoo, and James Won-Ki Hong are with the Department of Computer Science and Engineering, Pohang University of Science and Technology, Pohang 37673, South Korea ((e-mail: jhyoo78@postech.ac.kr, jwkhong@postech.ac.kr).

Ali Nauman is with the Department of Information and Communication Engineering, Yeungnam University, South Korea (e-mail: anauman@ynu.ac.kr).

Talha Ahmed Khan is with the Institute for Communication Systems, University of Surrey, Guildford, United Kingdom. (e-mail: t.a.khan@surrey.ac.uk).

Muhammad Afaq and Wang-Cheol Song are with the Department of Computer Engineering, Jeju National University, Jeju, South Korea (e-mail: afaq24@gmail.com, philo@jejunu.ac.kr).

Corresponding Author: Wang-Cheol Song

A. *Major Contribution*

Internet Engineering Task Force (IETF) introduced Intent-based Networking (IBN) for automatic and intelligent service provisioning, where users need to input intents, and the system handles all the operations for deploying the resources over the infrastructure [4]. In this work, we have proposed an innovative IBN platform with an AI-enabled network data analytics function for intelligent service provisioning and management, namely IBN. The proposed IBN architecture comprises an IBN system, NFV orchestrators, a RAN controller, real-time monitoring components, a data collection repository, and NWDAF, as presented in Figure 1. IBN performs intelligent service design and orchestration for efficient service provisioning and management for network slices. The primary contributions of IBN work are as follows.

- We have developed an IBN platform, an intelligent orchestrator for performing upper-level service design and orchestration tasks.
- IBN automates the service configuration generation by converting higher-level service requirements into network slice templates. IBN takes QoS requirements in the form of user intents for a slice as input and automatically converts them into a slice template for the core, mobile edge computing (MEC), transport, and RAN domains.
- To provide network intelligence that guarantees proactive lifecycle management of e2e network slices, we have integrated AI-enabled NWDAF with the IBN mechanism.
- IBN enables multi-tenants to automatically design, activate, monitor, and delete services.

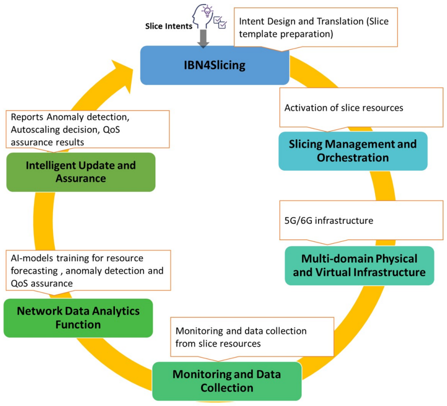

Fig. 1: Abstract design of IBN mechanism for automation and management of network slicing for 5G and beyond networks

The remaining paper is structured as follows. Section II contains the background of network slicing and management, IBN, and NWDAF for network automation. The innovative architecture of the proposed IBN for automatic network slicing is explained in Section III. Section IV presents the results of created network slices for services with the IBN mechanism. Lastly, the conclusion of the IBN work is summarized in Section V.

II. BACKGROUND TECHNOLOGIES

A. *Intent-based Networking System*

Many industrial and standardized organizations, such as the European Telecommunication Standard Institute (ETSI), IETF, Cisco, 3GPP, and next-generation mobile networks (NGMN), are developing intelligent frameworks to automate and manage future networks. These tools use AI technologies to achieve intelligence in the network and perform proactive resource management. For this purpose, IBN technology is the best choice for automating and managing networks. Cisco, Huawei, and APSTRA adopted IBN technology to automate and orchestrate future networks. IBN provides proactive control and management to fulfill customers' QoS requirements. It receives higher-level user intentions and automatically deploys them over the infrastructure. It is a self-configure, self-assured, self-organized, and self-healing mechanism. IBN consists of intent design, translation, activation, and intent assurance. Firstly, users need to input business intents through the system's dashboard. Secondly, received service requirements are translated into policies through the translation engine. Thirdly, translated policies are deployed over the physical and virtual infrastructure. Finally, deployed services are appropriately monitored to guarantee SLA [4]. Moreover, IBN enables AI algorithms to perform proactive intelligence in the network.

B. *Orchestration and Management of Network Slicing*

Automating and managing decentralized network slicing to support novel and diverse services is crucial for the upcoming 6G network. Many network standardization communities, including ETSI, IETF, and 3GPP, have developed network-slicing specifications. 3GPP has presented an architecture for e2e management and orchestration (M&O) of 5G mobile networks, which contains several management functions, such as communication service management function (CSMF), network slice management function (NSMF), and network slice sub-management (NSSMF). Besides, ETSI and IETF introduced zero-touch service management (ZSM) architecture to accommodate the 5G/6G services. Besides, several open-source NFVOs provide network slicing support, including OSM, Cloudify, open network automation platform (ONAP), open NFV (OPNFV), Tacker, SONATA, etc. [5]. Still, they have several limitations in automatically supporting network slice lifecycle management. These orchestrators require complex configurations to deploy the multi-domain resources for network slicing.

C. *Network Data Analytics Function*

The AI technologies enabled advanced network automation tools such as IBN, ZSM, and ONAP to perform data analytics based on historical network data and automatically control several networks' operations. Similarly, 3GPP presented NWDAF that collects data from core VNFs, RAN VNFs, physical network functions (PNF), and MEC VNFs to perform data analytics [6]. It allows the development of multiple AI methods trained on historical network data and performs several predictions and recommendations tasks such as network resource utilization prediction, mobility prediction, user behavior prediction, services recommendation, anomaly detection and prevention, autoscaling of slice resources, load balancing, service assurance, etc. Moreover, NWDAF assists MNOs in providing service guarantees to customers and managing resources proactively. So, NWDAF seems a perfect component for proactively managing multi-tier resources.

III. ARCHITECTURE OF PROPOSED FRAMEWORK

Figure 3 illustrates the main architecture of the IBN framework for proactive and intelligent service design and orchestration. It comprises an IBN framework, NFVO orchestrators, SDN-based RAN controllers, monitoring components, and multi-tier 5G/6G network infrastructure. IBN is designed as per 3GPP M&O architecture for mobile network slicing. IBN is an operation and support system (OSS) that provides the functionalities of the CSMF and NSMF management functions for service design and orchestration. The underlying NFVOs and controllers perform the NSSMF functionalities and manage the resource activation operations of each slice domain. In our implemented prototype, OSM (NFVO), FlexRAN controller, and SDN controller act as core, RAN, and transport NSSFM. In contrast, OSM handles core domain operations, FlexRAN oversees RAN domain operations, and transport-NSSFM manages transport network operations. So, the IBN platform automatically conducts the designing, activation, run-time monitoring, and deletion of network slicing for services. NWDAF assists IBN in proactively managing and updating multi-tier network resources. The working of each module of the IBN framework is well explained below.

A. Intent-based Networking for Service Orchestration

IBN is an intelligent system for autonomous orchestration and management of network slicing for a diverse range of services. IBN comprises two modules: intent design and translation and slice template generation. First, the intent design and translation module is responsible for getting user intents for a slice from the multi-tenants and automatically translating them to network policies. IBN has a web portal for multi-tenants to input their service requirements to activate a slice. These requirements are service type, uplink/downlink speed, slice single network slice selection assistant information (SNSSAI) ID. The E2E design and information repository is an IBN database repository that stores multi-tier network resources, service information, activated slice-intents, network topology information, and service function chaining (SFC) information for connecting core network VNFs. So, with the help of these two modules, the abstract slice-intents are translated into network policies and forwarded to the next slice template generation module. The slice template generation module contains platform-dependent slice template generators for the core, TN, and RAN domains. The core domain slice template generator is OSM-dependent and prepares slice templates by inputting received network policies from the intent design and translation module. On the other side, based on the received policies, the RAN template generator designed a slice template for RAN slicing. Furthermore, these slice templates are forwarded to NFVOs and RAN controllers to activate multi-tier resources. So, the first module converts higher-level slice-intents into network policies, and the second module prepares slice templates for underlying domains. All the components communicate using REST interfaces [5], [7]. IBN system used the following components to perform network slicing over the multi-tier infrastructure.

B. Slice Resource Activation and Management

1) Core Network Management: OSM is the most popular ETSI MANO-based NFVO for the deployment of VNFs. It provides NFVO, VNF manager (VNFM), and virtual infrastructure manager (VIM) capabilities for automating and managing network instances [8]. OSM provides an integrated OpenStack platform as a VIM for deploying core network VNFs. So, in our IBN mechanism, we have used OSM as an NFVO for handling the deployment of core VNF instances. So, the core slice template generator designed the slice template in JSON-string for OSM. OSM accepts policies in JSON string format, such as Network Service Descriptor (NSD), Virtual Network Function Descriptor (VNFD), and Network Slice Template (NST) through REST interfaces for the deployment of resources. NSD has information about how a deployed VNF will connect to provide a service. In addition, VNFD contains information related to interfacing, networking, and resources. IBN pushed the prepared slice template to OSM through REST-API for deploying core network instances.

2) RAN Management: To handle RAN domain operations for performing network slicing, we have used an SDN-based RAN controller, FlexRAN. FlexRAN emerged as an intelligent controller because of its openness to develop multiple applications on top of it for handling many RAN operations [9]. It also supports dynamic RAN slicing. So, in our mechanism, IBN pushed the JSON-based slice template prepared by the RAN slice template generator module through the REST interface to FlexRAN to slice the RAN resources as per QoS.

3) Run-time Monitoring and Data Collection: Monitoring underlying multi-tier resources is crucial for closed-loop automation and management. So, in our mechanism, we have used widely adopted monitoring tools such as ElasticMon, Grafana, and Prometheus to monitor multi-domain virtual and physical resources and collect the data from the underlying resources. Grafana is a visualization tool that can visualize the traffic and resource usage for better management, and Prometheus is a tool to monitor the resources in a time series manner. It supports monitoring cloud-based infrastructure and helps to store the logs in time-series databases for future use [5]. The monitoring module helps to store time-series logs and events from the activated slice resources. We used various node exporters to collect and store logs on a server. These logs are collected from the core, MEC, and RAN resources.

4) NWDAF for Service Update and Assuranc: NWADAF is a module to enable network intelligence to the IBN framework. We can develop several AI models for performing different

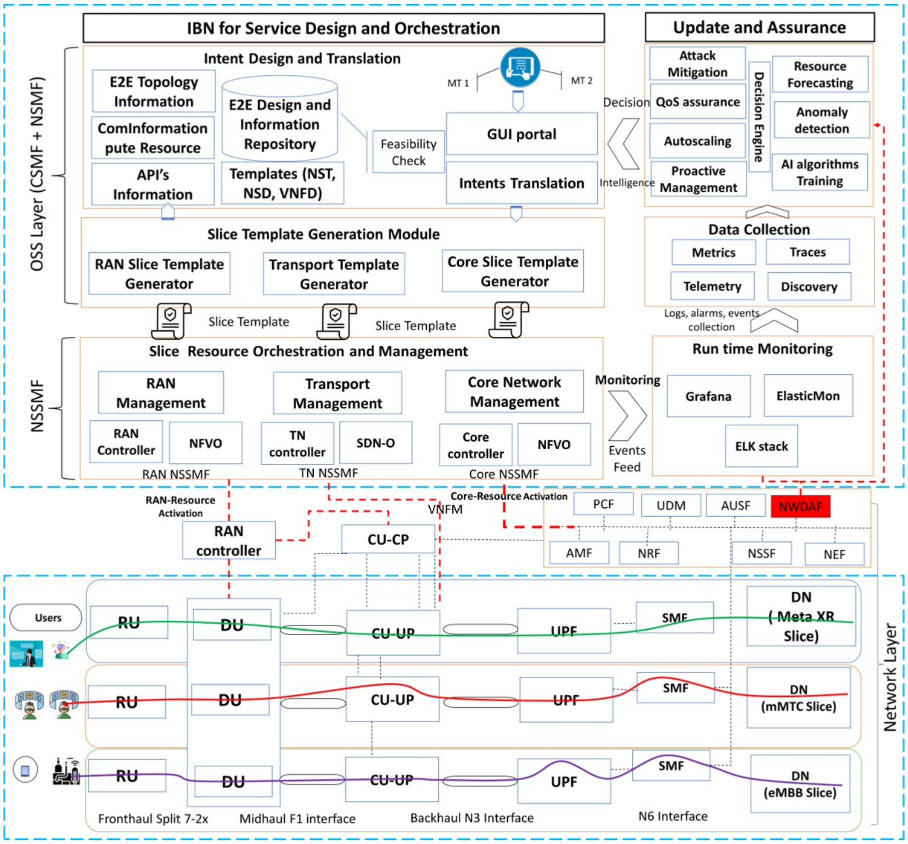

Fig. 2: Illustrates the detailed architecture of the IBN framework for e2e service automation and management. It also illustrates the components of the testbed, including 5G core and RAN, orchestrators, controllers, monitoring of resources, data collection, and AI-based NWDAF for update and assurance.

tasks inside the NWDAF. These AI methods enable MNOs to perform proactive management tasks such as core and RAN resource usage prediction, anomaly detection, autoscaling, load balancing of VNFs, service recommendation, mobility prediction, and failure prediction. We've strategically deployed multiple node exporters within each domain to facilitate real-time data collection from the underlying Core and RAN resources. This ensures a constant stream of up-to-date data. Subsequently, the collected data is stored in real-time NWDAF databases, where it's readily accessible. With the data in hand, we employ various ML techniques for meticulous data preprocessing. This includes essential steps such as data cleaning, transformation, feature engineering, and feature reduction, all geared toward optimizing data quality and utility. Once the data is suitably prepared, it is fed into multiple ML models for comprehensive training. These ML models undergo rigorous evaluation, focusing on key performance metrics like accuracy and root mean square error (RMSE). After successful evaluation, the ML models are implemented for real-time prediction or classification tasks. For instance, in VNFs resource forecasting cases, ML models such as time series LSTM and hybrid ML models provide invaluable predictions to the IBN, which, in turn, formulates policies to address resource overloading and underloading issues. Similarly, in the realm of network attack detection, the ML models promptly identify potential threats and communicate their findings to the IBN. Armed with this information, the IBN swiftly initiates mitigation policies to counteract and halt the detected attacks. The NWDAF provides alerts to IBN using the results of ML methods, and IBN can update the resources before the SLA (service level agreement) violation and service degradation. This intricate process exemplifies the dynamic synergy between AI and the IBN mechanism, ensuring the network operates efficiently and securely.

C. Multi-Domain Infrastructure

The last layer is a multi-domain 5G/6G network infrastructure with enough capabilities to support novel service requirements. In our implemented prototype, we have used open-source implementation of 5G core VNFs: access and mobility management function (AMF), user plane function (UPF), session management function (SMF), and gNB for RAN from the OpenAirInterface (OAI) [10]. As per the 5G slicing architecture, each service should be provided with dedicated and shared 5G gNB (radio unit (RU), centralized unit (CU), distributed unit (DU)) wireless resources. Each service is provisioned with a CU user-plane (CU-UP) and

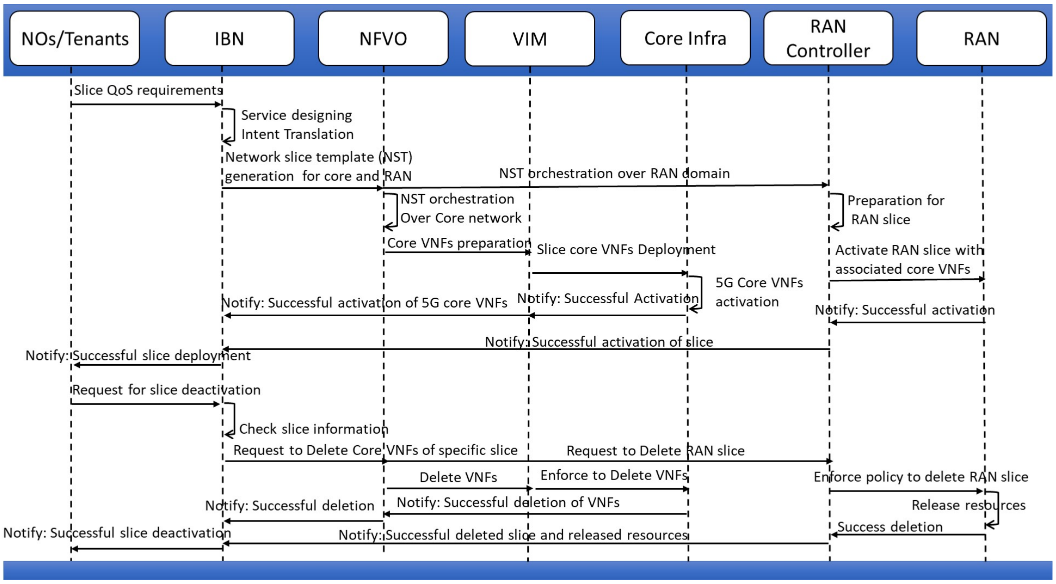

Fig. 3: illustrates the slice designing, activation, and deactivation workflow using the IBN mechanism. It also depicts the communications of all the components, including IBN, NFVO, VIM, and RAN controller of the system involved in automated activation and deletion of e2e network slicing.

a dedicated UPF for guaranteeing QoS. So, network slicing enables multi-tenants to provide customized resources to their users by guaranteeing QoS with better QoE.

Figure 3 depicts the slice designing and activation process through our IBN mechanism. Specifically, the IBN is a platform for automated network slice development (NST) development in which users need to enter service QoS through the web portal. The IBN manager checks slice QoS requirements from the database repository, which has all the information related to network topology, computer resources, slice-id, and slice type. After getting the required information from the database repository, the intent manager forwards specific details to the slice template generation module. The IBN has separate slice template generators for each domain, such as a core, TN, and RAN. The slice template generator receives the information from the intent design and translation module and converts it into policies of specific formats such as YAML and JSON. Each slice template generator has a built-in template for a specific underlying platform. It inserts the service QoS information from the upper module into that template. After completing the slice templates, template generators forwarded them to the OSM and FlexRAN to deploy resources for services. OSM with OpenStack deploys the VNF instances for the core slice; the FlexRAN controller deploys the RAN resources for establishing the RAN slice, and the SDN controller handles TN operations. The procedure for deactivating and releasing deployed slice resources commences with initiating a slice deletion request via the IBN portal, as presented in Figure 3. The IBN then relays this request to the underlying entities, including the NFVO and RAN controller, along with essential slice information. The NFVO, working with the VIM, enforces policies designed to delete the core VNFs. Concurrently, the RAN controller sets the RAN resource deletion policy in motion, releasing RAN resources. After completing these deletion processes, notifications are promptly disseminated to the stakeholders through the IBN, ensuring that all concerned parties are duly informed of the slice deactivation and resource release.

IV. PROTOTYPE IMPLEMENTATION AND RESULTS

The Implemented network slicing prototype comprises an IBN, FlexRAN controller, OSM platform, NWDAF, and OAI components: 5G core VNFs, gNB, and UEs. The OAI provides the implementation of RAN and 5G core VNFs. Conversely, the FlexRAN controller deployed and controlled RAN slicing. The gNB was used as a RAN in our testbed with software-defined radio (SDR) universal software radio peripheral (USRP) B210. The USRP B210 is connected to gNB and provides radio capabilities for UEs. The OSM and OpenStack are deployed in a server machine with 32 cores and 252GB memory. IBN application has been developed in Python, Java, and MYSQL language. The database repository of IBN has all the information about the infrastructure and deployed slices and slice templates for RAN and core domains. The deployed network topology is based on 5G architecture, which contains UE, gNB, and core VNFs such as AMF, UDM, SMF, and UPF. So, the core VNFs and RAN resources are chained together using 5G topology to provide e2e connectivity. The traffic is generated from UE through the iPerf test to validate system connectivity. Every slice has a dedicated core network AMF function for e2e connection.

To test the stability of our IBN mechanism for network slicing, we have performed iPerf tests. We have deployed multiple slices with different service QoS requirements. Figure

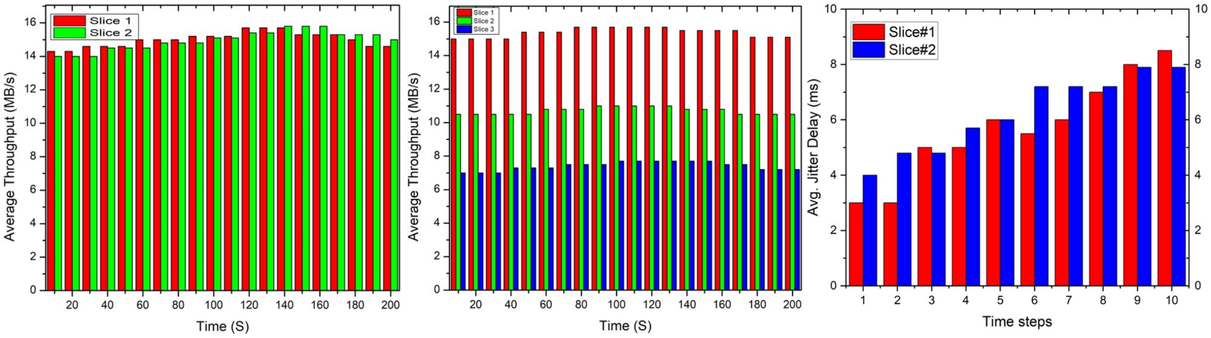

Fig. 4: illustrates iPerf results of the activated slices through IBN platform a) two network slices were activated with similar QoS requirements b) three slices were activated with different QoS requirements c) presents the calculated average jitter of two activated slices from multiple tests

4 shows two deployed slices, namely slice 1 and slice 2, with 50% of RAN resources to each; by inserting service QoS requirements through the IBN web portal, the system automatically activates the network slice with the cooperation of other components. We are limited with RAN resources because we have used USRP SDR for RAN. However, we considered the maximum RAN resources for activating the e2e slices. Figure 4 a) shows the downlink speed recorded through the deployed e2e slice 1 and 2. Slice 1 and slice 2 slice show almost similar throughput maximum of 15.8 MB/s because equal resources are requested through the slice intents. Figure 4 b) presents the results of three slices deployed with different user intents through the IBN platform. Slice 1, slice 2, and slice 3 were deployed with 50%, 30%, and 20% of the RAN resources. So, the maximum of 15.5 MB/s, 11 MB/s, and 7.8 MB/s downlink speed were recorded for slice 1, slice 2, and slice 3, respectively. Figure 4 c) illustrates the results of the achieved average packet jitter for two deployed slices. Slices 1 and 2 show a maximum of 8ms and 7ms jitter delay, respectively, calculated from multiple tests. So, these results show the stable performance of deployed slices for serving various types of users with specified QoS requirements. The experimental outcomes of activated slices reveal that our IBN mechanism is efficient and stable while automating network slicing. With many advantages, there are a few limitations to our IBN mechanism. Firstly, the efficient and dynamic allocation of RAN resources still needs to be improved. We plan to develop a reinforcement learning-based application to dynamically provision RAN resources for each slice type. Secondly, our mechanism still needs to extend IBN to achieve zero-touch network automation. Thirdly, we have to enhance our work by developing automated Blockchain and AI-supported network security mechanisms for securing 5G and beyond networks.

V. CONCLUSION

Future networks can accommodate an extensive range of innovative services with distinct QoS requirements regarding latency and bandwidth. This work explained the IBN platform for the autonomous orchestration of network slices. IBN framework automates the complex network configuration generation process and provides an efficient and error-prone way to generate slice templates for underlying resources. IBN framework automatically designs, activates, monitors, and deactivates the network slices. In addition, multiple tests have been performed by creating several slices of core and RAN domain through our mechanism, showing satisfactory results in resource stability, automated resource provisioning, customization, and resource assurance. Conversely, AI-enabled NWDAF with IBN ensures proactive updates and assurance for services. We will extend the IBN mechanism in the future by developing reinforcement learning-based admission, management, and control of network slices. We will also enhance our mechanism by providing a zero-trust security mechanism to ensure network security.

REFERENCES

[1] F. Tang, X. Chen, M. Zhao, and N. Kato, "The roadmap of communication and networking in 6g for the metaverse," *IEEE Wireless Communications*, 2022.

[2] I. Afolabi, T. Taleb, P. A. Frangoudis, M. Bagaa, and A. Ksentini, "Network slicing-based customization of 5g mobile services," *IEEE Network*, vol. 33, no. 5, pp. 134–141, 2019.

[3] K. Katsalis, N. Nikaein, E. Schiller, A. Ksentini, and T. Braun, "Network slices toward 5g communications: Slicing the lte network," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 146–154, 2017.

[4] IETF, "Intent-based networking," https://tools.ietf.org/html/draft-irtf-nmrg-ibn-concepts-definitions-01.

[5] K. Abbas, T. A. Khan, M. Afaq, and W.-C. Song, "Network slice lifecycle management for 5g mobile networks: An intent-based networking approach," *IEEE Access*, 2021.

[6] E. Pateromichelakis, F. Moggio, C. Mannweiler, P. Arnold, M. Shariat, M. Einhaus, Q. Wei, Ö. Bulakci, and A. De Domenico, "End-to-end data analytics framework for 5g architecture," *IEEE Access*, vol. 7, pp. 40 295–40 312, 2019.

[7] K. Abbas, M. Afaq, T. Ahmed Khan, A. Rafiq, and W.-C. Song, "Slicing the core network and radio access network domains through intent-based networking for 5g networks," *Electronics*, vol. 9, no. 10, p. 1710, 2020.

[8] OSM, "Open source mano," https://osm-download.etsi.org/ftp/Documentation/201902-osm-scope-white-paper/#!02-osm-scope-and-functionality.md;accessed:1.6.2020.

[9] X. Foukas, N. Nikaein, M. M. Kassem, M. K. Marina, and K. Kontovasilis, "Flexran: A flexible and programmable platform for software-defined radio access networks," in *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*, 2016, pp. 427–441.

[10] OpenAirInterface, "Oai: An open-source community," https://www.openairinterface.org/;accessed:10.5.2020.