5G-Serv: Decoupling User Control and Network Control in the 3GPP 5G Network

Meghna Khaturia*, Akshatha Nayak Manjeshwar*, Pranav Jha* and Abhay Karandikar*†
*Department of Electrical Engineering, Indian Institute of Technology Bombay, Mumbai, India
†Director, Indian Institute of Technology, Kanpur, India

***Abstract*—The Third Generation Partnership Project (3GPP) Fifth Generation (5G) network employs Software Defined Networking (SDN) paradigm allowing for a clear separation of control and data plane functionalities in the 5G network. However, the control plane in a 5G network not only performs network control tasks involving control and management of data plane functions but is also responsible for executing UE-specific control tasks. We think these two tasks are independent of each other and propose to decouple them and present 5G-Serv, a new architecture for 5G mobile network. The paper presents preliminary analysis and evaluation of the proposed architecture through flow diagrams. As demonstrated, 5G-Serv simplifies the information flow in the 5G network and enhances its modularity and flexibility. The architecture may also be better aligned with the standard SDN paradigm vis-a-vis the 3GPP architecture.**

I. Introduction

5G and beyond mobile network is being designed to support billions of connected devices/users and complex use cases such as enhanced Mobile Broadband (eMBB), ultra-Low Latency Communication (uRLLC), Massive Machine Type Communication (MMTC). To meet these requirements and enhance the network's flexibility and programmability, adoption of Software-Defined Networking (SDN) has been one of the key advancements in the Third Generation Partnership Project (3GPP) 5G standardization. Following the SDN paradigm, there exists a clear split of control and data plane functionalities, both within the 5G Core Network (CN) and the Radio Access Network (RAN).

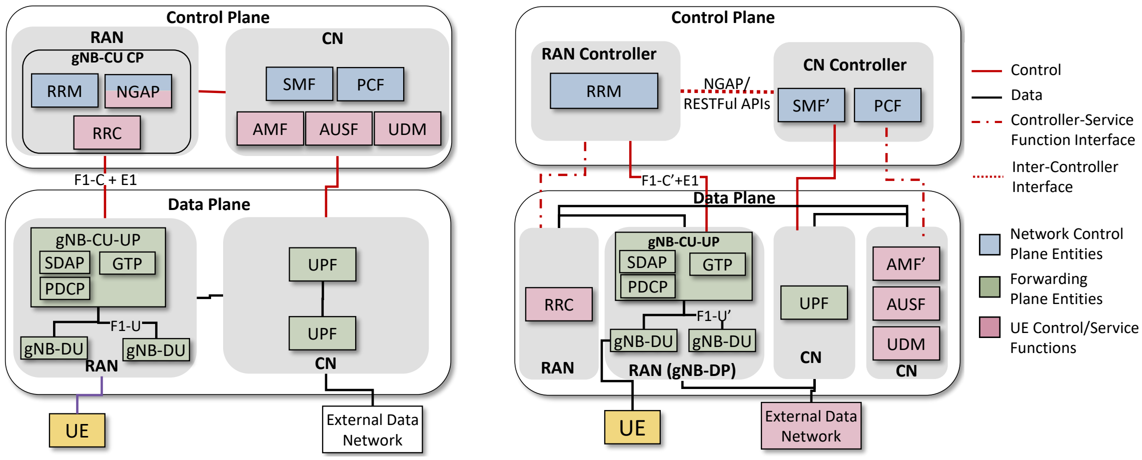
On closer observation, it is apparent that the tasks performed by Control Plane (CP) functions within the 3GPP 5G network can further be classified into tasks related to, i) handling signaling message exchange with User Equipment (UE) (UE control and management) and ii) those pertaining to control and management of Data Plane (DP) functions in the network. CN CP functions such as Access and Mobility Function (AMF) and Session Modification Function (SMF) communicate with UEs using Non-Access Stratum (NAS) layer protocol and collect data transfer related requirements from them. At the same time, they are also responsible for network control tasks, such as establishment and modification of Protocol Data Unit (PDU) sessions over the DP functions (i.e., User Plane Function (UPF)). Similarly, the CP function in RAN (gNB) exchanges signaling messages with UEs via Radio Resource Control (RRC) protocol and also controls DP functions in RAN (gNB). Note that the 3GPP 5G architecture does not distinguish between these two categories of tasks and considers both to be a part of the CP functionality. However, in our view, these are independent and can be decoupled from one another, as has been proposed in this paper.

In a traditional SDN based network architecture [1], an SDN Controller (network control function) performs tasks to enable data transfer through the network. Tasks such as establishing or modifying data paths/flows over forwarding plane functions fall in this category. However, network control functions are typically not responsible for communication with end-users to gather the requirements from them. Our proposal to split the existing 5G CP into two sets of functions, one responsible for the exchange of information with UEs and the other one responsible for control and management of the DP, is aligned with this view of the SDN based architecture proposed by IRTF/IETF standards [1], [2]. Works such as [2]–[4] have also emphasized the need for a well-defined interface between UE control and network control interaction. Due to the absence of this interface, network management in the existing 3GPP 5G network is complex [3]. To the best of our knowledge, these ideas have not been sufficiently explored in existing literature, and none of the available literature provides an end-to-end network architecture aligned with this objective.

In this paper, we examine how 3GPP 5G network architecture transforms if we decouple UE control from network control functionality. We propose a new architectural framework "5G-Serv", to decouple these two functions and reorganize the CP of the 5G mobile network in order to improve its modularity and flexibility. Our previous work in [5] briefly explored this idea for 5G RAN. In summary, the major contributions of the paper are —

- We propose enhancements to 3GPP 5G architecture to make the network more modular. This is achieved by separating UE control functionality from network control functionality wherein a set of new functions called service functions is responsible for communicating with the UE and gathering requirements while the network controller is responsible for control and configuration of the data plane functions.
- We explain the working of the proposed architecture and illustrate a few procedures within the proposed architecture with the help of flow diagrams.
- The work is in an early stage and therefore detailed evaluation of the proposed architecture is not yet available. However, reduction in UE signalling due to the modular nature of the architecture is demonstrated through call flows.

The rest of the paper is organized as follows. In Section II, we provide an overview of the 3GPP 5G network and also describe the proposed network architecture. The working of proposed architecture is described in Section III. We present some advantages and conclude our work in Sections IV and V, respectively.



(a) Existing 3GPP 5G Network Architecture (b) Proposed 5G-Serv Architecture

Fig. 1: Proposed 5G-Serv Architecture

II. Proposed Architecture

A. *Description of the Existing 3GPP 5G Network*

As shown in Fig. 1a, 3GPP 5G network comprises DP and CP functions. Here we discuss a few important functions of 5G network. In the 5G CN, UPF performs the role of the DP function, forwarding user data through the CN. AMF enables UE registration, mobility management, and UE's connection with the CN. It also handles NAS message termination for UEs. AMF performs a few network control tasks such as relaying session management related messages to SMF. SMF handles network control tasks, such as creating/modifying and releasing PDU sessions and management of UPF. However, SMF also performs a few UE control tasks, such as NAS message handling for Session Management and IP address allocation and management for UEs. PCF enforces network policy to govern network behavior, and AUSF serves as an authentication server. UDM hosts a user database that can be used by other network functions.

The 3GPP 5G RAN comprises base stations known as gNBs. A gNB is constituted by a gNB-Centralized Unit (gNB-CU) and one or more gNB-Distributed Units (gNB-DUs). A gNB-CU further comprises CP and UP parts, i.e., gNB-CU-CP and gNB-CU-UP. The gNB-CU-CP provides CP termination towards a UE. It comprises RRC, Radio Resource Management (RRM), and Next-Generation Application Protocol (NGAP). The RRC facilitates radio connection establishment and release for a UE. RRM is responsible for radio resource allocation for the 5G RAN. RRM also handles UE connection and mobility management related functionality such as handover management and admission control. NGAP facilitates UE control tasks such as transporting the UE NAS messages between gNB and CN. The network control tasks of NGAP include transportation of session and mobility management messages between gNB and CN. Besides, gNB-CU-CP is also responsible for controlling and managing RAN DP functions, gNB-CU-UP, and gNB-DU through E1AP and F1AP protocols, respectively. gNB-CU-UP comprises Service Data Adaptation Protocol (SDAP), GPRS Tunnelling Protocol (GTP), and Packet Data Convergence Protocol (PDCP). SDAP is responsible for mapping the QoS to the radio bearers. GTP, along with its underlying stack, is used to exchange data with CN. PDCP and the underlying protocol layers are responsible for transmission/reception of data over the radio interface.

B. *Network-Service Model*

We introduce a generalized SDN based network service model (shown in Fig. 2) to better explain the 5G-Serv architecture.

The primary function of a communication network is to enable the exchange of information (data) between two end-points. An example is an exchange between a UE and a 'content server' where the server provides audio or video content to the UE. Such an exchange of information can be called a 'Service'. Another example may be access control and authentication service in mobile networks where an entity in the network authenticates a UE, and further access is granted to it. Providing these services may typically need deployment of dedicated functions in the network, e.g., a content server, AUSF (in 3GPP 5G network) as mentioned above. In this paper, we propose to use the term 'Service Function' to denote these dedicated network functions.

The facilitation of a service to a UE may involve two steps a) the collection of service requirements and b) the establishment of a path for service access. Interestingly, even collecting service-related requirements from a user may need establishment of a path through the network. For example, to establish a data path (a PDU Session to connect to Internet) in the 5G network, the requirements for the path is provided by the UE to 5G AMF/SMF functions via NAS signaling messages. However, to enable the exchange of NAS messages itself, a dedicated path, i.e., a signaling radio bearer (SRB) between UE and gNB, needs to be established first. This SRB is established by the RRC layer. Therefore, we can view the network operation as a feedback loop, i.e., it is recursive in nature. In this recursive view, the network control functions (SDN Controller(s)) enable communication between a UE and a group of service functions. These service functions may trigger the network CP functions using a separate module,

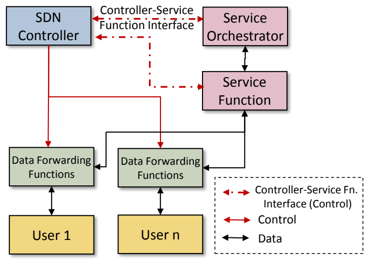


Fig. 2: Network Service Model

called a 'Service Orchestrator' to enable additional services for the UE through establishment of supplementary data paths. Service Orchestrators, shown in Fig. 2, are responsible for orchestration of a service with the help of the controllers.

C. Proposed 5G-Serv Network Overview

Harmonizing the flow of information in 5G network with the network service model defined in Section II-B, we propose '5G-Serv', a novel SDN based architecture for 5G and beyond mobile network. The proposed architecture comprises (a) Data plane functions, responsible for data forwarding through the mobile network (b) Network control functions, responsible for control and management of DP functions, i.e., establishment/modification of data paths and (c) Service functions, responsible for information exchange with UEs and triggering network control functions to establish data paths.

An important implication of building the network architecture on the proposed network service model is that, the service functions, even the ones responsible for handling UE signaling messages (UE control functions), e.g., gNB RRC Layer or AMF can be considered as part of DP as they are responsible for exchange of information with UEs. This is similar to the exchange of information between UE and a service function such as a content server. Controller may view the UE specific signaling messages such as RRC messages or NAS messages as a type of data only and establishes paths for them through the DP in the network, i.e., through gNB-DU and/or gNB-CU-UP etc. This also means that the controller and the network do not distinguish between UE signaling messages (carried over SRBs) and UE data (carried over PDU Sessions/DRBs) and treat both of them similarly.

Now, we introduce modifications in the 3GPP 5G network to realize 5G-Serv, as shown in Fig. 1b. As indicated earlier, the primary goal of 5G-Serv is to separate UE control function from network control function. We observe that the RRM module in existing gNB-CU-CP is typically responsible for resource allocation and control of DP functions in RAN, while the RRC layer is used for signaling exchange with UEs over the radio interface. Therefore, we transpose the RRC functionality from the gNB-CU-CP into a service function called 'RRC'. The new RRC function is located outside the gNB-CU-CP, and the signaling message exchange between the RRC function and the UE takes place directly via RAN DP without involving gNB-CU-CP. The movement of RRC out of gNB-CU-CP simplifies the F1AP protocol as it (F1AP) is no longer required to carry RRC messages between gNB-CU-CP and gNB-DU. This simplified interface is represented as F1' in the figure. Along with RRC functionality, handling of UE-specific NGAP messages (e.g. the ones carrying NAS messages between AMF and gNB-CU-CP in the existing 5G network) is also moved out of gNB-CU-CP as it is no longer in the path of NAS message exchange between UE and CN. With a considerably reduced scope, the resulting gNB-CU-CP called the 'RAN Controller' is responsible for handling only network control functionality. 5G-Serv provides an inter Controller interface viz., the *RAN-CN Interface*, wherein the RAN and CN controllers can communicate with each other. This interface can be defined using a subset of NGAP messages (messages related to data session setup/modification/release, etc.) or RestFul APIs.

Similar to RAN, we decouple the UE control tasks from network control tasks in CN. We introduce a simplified AMF (AMF') in 5G-Serv CN. AMF' acts as the endpoint for all NAS message exchange with UEs and performs the UE Control functionality. We also modify the 3GPP SMF (SMF'). SMF' is a simplified version of the 3GPP SMF, wherein the NAS message handling has been removed completely. SMF' is responsible for resource allocation in CN and is a part of the CN Controller, a new CN CP entity in the proposed 5G-Serv architecture. In addition, PCF is also a part of the CN CP entity in 5G-Serv. AMF' connects to RRC directly to send/receive NAS messages. AMF' interprets NAS messages received from UEs and understands their requirements. Once it receives a UE's requirements, such as requesting a new PDU Session, it requests SMF' to set up the required data session. Here AMF' can be viewed as a service function in CN just like 'RRC' in RAN. It is handling NAS service for the UE. It also acts as a service orchestrator for CN Controller since it sends triggers for resource allocation to the Controller. We also place AUSF and UDM as service functions alongside AMF' in 5G-Serv CN as they are responsible for supporting a service like 'authentication and access control' for a UE.

The data forwarding functions (gNB-DU, gNB-CU-UP and UPF) are responsible for processing and routing data packets over the network. The DP of the proposed network remains same as that of the 3GPP 5G network. For brevity, we refer to the DP functions in RAN (gNB-DU and gNB-CU-UP) as gNB Data Plane (gNB-DP) in the rest of the paper.

To forward the service related requirements between AMF' and CN Controller or between RRC and RAN Controller, 5G-Serv introduces a *Controller-Service function interface*. This interface can also be used by other service functions to communicate with the Controller and send request for resource allocation in the network. With the introduction of this interface, the service functions (e.g., RRC, AMF etc.) can be placed anywhere in the network. This advantage is explained in Section IV.

III. UE SERVICES IN 5G-SERV

In this section, we demonstrate a few example procedures with the help of information flow diagrams.

A. PDU Session Establishment

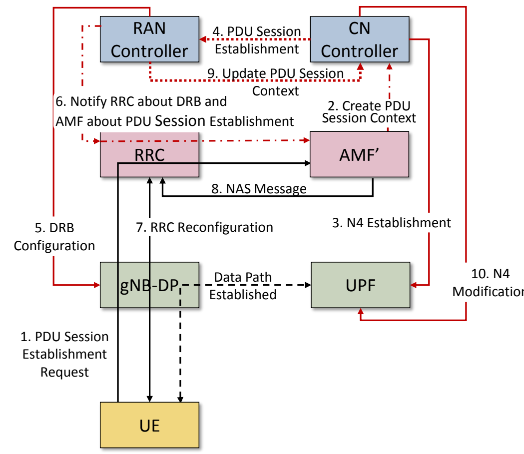
A UE accesses *Registration* service to establish communication with the 5G CN. Once the Registration process is complete, a UE can request additional services via the default (signaling) path established during the *Registration*. This aligns with the recursive view of the proposed network. For example, within 5G-Serv, a UE can access data services through *PDU Session Establishment* Procedure, as shown in Fig. 3a. To initiate this, a UE sends a *PDU Session Establishment Request* message to the UE Control Function (to AMF'). To facilitate this service, AMF' forwards this request to the CN controller. The SMF' within the CN controller configures the UPF to set up a GTP tunnel for the PDU session. On the proposed RAN-CN controller interface, the CN controller informs RAN Controller about PDU session related attributes (GTP tunnel ID) for setting up resources for the PDU session in RAN. Subsequently, the RAN controller creates a GTP tunnel at the gNB and configures DRB for the PDU session. Once the resource setup is complete, the RAN controller notifies RRC, and CN Controller notifies AMF'. RRC initiates *RRC Reconfiguration* to set up the DRB and also inform UE about the PDU session establishment. The end-to-end resource allocation for the data service is now complete. Now, the RAN controller updates the CN controller regarding the successful completion of the PDU session establishment. CN controller then provides the GTP tunnel info to the UPF.

As explained earlier, interaction between UE and service functions can be treated as data. Therefore, in the figure, all the messages exchanged between UE and service functions are shown as data. Any communication with the controllers is considered control communication. The decoupling of service functions and network control is apparent from Fig. 3a. The AMF (service function and service orchestrator) communicates with SMF for orchestrating network services, whereas SMF (within CN Controller) configures the network to enable these services. This also simplifies the inter-working between network functions in the network and also enables us to establish well-defined interfaces between them.

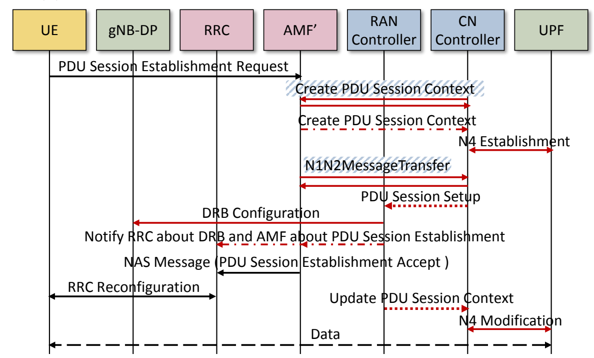
As shown in Fig. 3b, the PDU Session Establishment in 5G-Serv is compared with 3GPP 5G PDU Session Establishment [6] and we observe that the procedure is simplified in the following manner. As observable, signaling messages relating to session setup and modification are exchanged between RAN and SMF. These messages are transmitted transparently through the AMF due to the absence of a direct path in 3GPP 5G network. However within 5G-Serv, due to the newly introduced RAN-CN interface, SMF can directly communicate with RAN Controller, eliminating a hop between them. Moreover, a hop is reduced for all the SMF related messages that are transparently exchanged through the AMF [7]. Note that it is possible to further simplify the signaling for the above procedures by co-locating the RRC and AMF.

B. Mobility Management

In mobile networks, mobility can also be thought of as a service, wherein the network to serve data to a user while it (user) moves. The 3GPP 5G network supports multiple



(a) Information Flow for PDU session establishment



(b) PDU Session Establishment call flow in 5G-Serv (shaded messages indicate eliminated 3GPP messages)

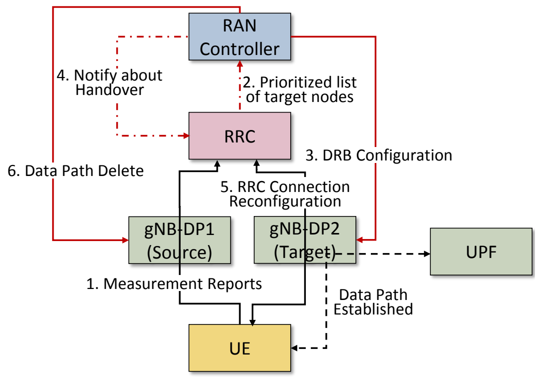
Fig. 3: PDU Session Establishment

variants of mobility. In this section, we consider an example of intra-gNB mobility within our framework. The procedure for intra-gNB mobility wherein a UE is handed over between two gNB-DUs which are under the control of a single gNB-CU is illustrated in Fig. 4a. UE sends periodic radio measurements to RRC, which decides whether a handover is required. If handover is needed, it calculates a list of suitable target gNB-DPs and organizes them on the basis of their priority. This list is then forwarded to the RAN controller. The RAN controller then selects a target node based on traffic load and available radio resources at the gNB-DP nodes. It sets up a radio bearer at the target gNB-DP to facilitate UE data transfer and then informs RRC to communicate to the UE about the DRB. UE uses the information provided in the RRC message to setup the DRB on its side. Once the new path is established between the target gNB-DP and UE, the handover is complete. The RAN controller now deallocates the radio resources reserved for the UE on the source gNB-DP. From Fig. 4b, we can observe that the request-response messages required for resource setup within the DP are replaced by a single configuration message. The RAN Controller provides these configurations over the F1' interface. F1' supports a simplified version of F1AP (excluding request-response messages).

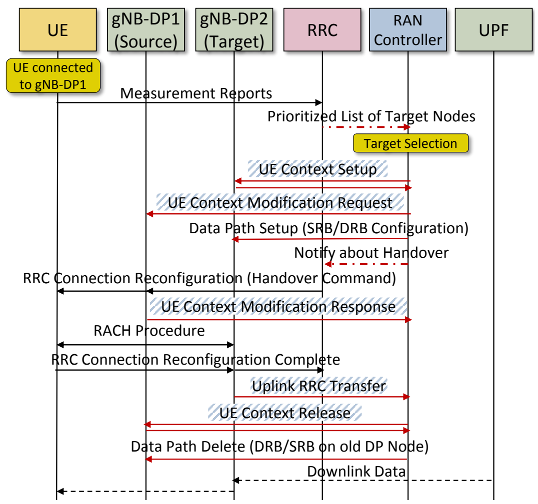
IV. Advantages

Some of the advantages of the 5G-Serv network are:

- *Simplified Network Architecture*: By decoupling service



(a) Information Flow for Mobility



(b) Mobility call flow in 5G-Serv (shaded messages indicate eliminated 3GPP messages)

Fig. 4: Mobility Managment

functions and network control functionalities, we are able to realize a modular network. Additionally, network functions such as AMF, SMF and interfaces such as F1 are simplified.

• *Facilitates Use of Simpler Protocols*: 5G-Serv provides simpler protocols with better separation of concern. For example, in 3GPP 5G network, F1AP is used to configure gNB-DU and also used to carry RRC messages. In 5G-Serv, F1AP is only used to configure the RAN data plane function i.e, gNB-DP. Due to this separation of concern, network operation as well as service enablement can be viewed as a set of recursive rudimentary tasks comprising path setup and modification.

• *Ease of Introduction of New Services:* 5G-Serv provides a well-defined interface for communication between service functions and control functions in the network, i.e., the Controller-Service function interface. With the help of this interface, modification of existing services and the introduction of newer services in the network becomes easier as the application developer is aware of the network capabilities.

•*Flexible Service Function Placement and Chaining:* As mentioned earlier, service functions can be flexibly placed in the RAN or the CN based on the network/service requirements. In 5G-Serv, a UE has the flexibility to choose services offered by the network. For instance, a user can choose a simpler authentication for standard data services, which reduces the signaling messages required for two-way authentication. Hence, the load at CN also reduces. This in turn supports increased scalability.

• *Improved Load management in RAN CP*: Movement of RRC layer functionality out of gNB-CU-CP helps in load distribution of RAN CP across multiple logical functions. We can have multiple RRC servers to handle RRC signaling messages from the UE and the RAN Controller can handle a large number of data plane entities.

• *Enable use case specific variants of UE signaling Protocols and associated state machines*: URLLC and eMBB may have different versions of RRC protocol and state machine and separate/different RRC servers can easily be used to support them. It may not be easy to support multiple variants of RRC in a single gNB-CU-CP but with independent and dedicated RRC servers it is easy to support these variants.

V. Conclusion and Discussions

In this paper, we have proposed '5G-Serv', an architecture for decoupling 'UE control' functionality from the 'network control' functionality within the 5G network. It brings a clear separation of concern to 5G CP functions, which is currently lacking in the existing 5G network. The architecture generalizes the concept of 'service', wherein all information exchange with UE is treated as 'service' (and thereby as a form of data), allowing us to include even exchange of signaling messages, between a UE and the network, within the ambit of 'service' (or data). It also enables us to view the deployment and modification of services in the network as a recursive process bringing inherent simplicity and enhanced modularity and flexibility to the 5G mobile network architecture. The architecture standardizes (unifies) the process of service provisioning in 5G mobile network making it easier for it to cater to existing and future service requirements. A plausible extension of this work is to demonstrate the performance improvements quantitatively through a prototype and to standardize the Controller-Service function Interface.

VI. Acknowledgement

This work has been supported by the Department of Telecommunications, Ministry of Communications, Government of India as part of the indigenous 5G Test Bed project.

References

[1] E. Haleplidis, K. Pentikousis, S. Denazis, J. Salim, D. Meyer, and O. Koufopavlou, "RFC 7426: Software-Defined Networking (SDN): layers and architecture terminology," *IRTF*, 2015.

[2] Q. Wu, W. Liu, and A. Farrel, "Service Models Explained ," RFC 8309, January 2018, Tech. Rep.

[3] A. Clemm, M. F. Zhani, and R. Boutaba, "Network management 2030: Operations and control of network 2030 services," *Journal of Network and Systems Management*, pp. 1–30, 2020.

[4] "ONF TR-521," Tech. Rep. 1.1, 2016.

[5] P. K. Taksande, P. Jha, A. Karandikar, and P. Chaporkar, "Open5G: A Software-Defined Networking Protocol for 5G Multi-RAT Wireless Networks," in *WCNC*, 2020.

[6] "3GPP TS 23.501, System Architecture for the 5G System," Tech. Rep., 2018, (Release 15).

[7] "3GPP TS 38.413, NG-RAN; NG Application Protocol (NGAP)," Tech. Rep., 2020, (Release 16).