A Bottom-up Network Slicing Procedure in F-RAN

Tao Wang¹*, Zhaopeng Zhu¹ and Yu Hu¹

¹ School of Electronic and Information Engineering, Xi’an Jiaotong University, Xi’an, China

*Corresponding author e-mail: wongtao@stu.xjtu.edu.cn

Abstract. As a core technology of future mobile communication networks, network slicing can provide users with end-to-end, customized, and isolated network connections. Thereby it is of great significance for improving network resource utilization and improving user experience. F-RAN, as a solution for 5G access network, has features such as low latency, support of mobility, and content awareness. Through the analysis of traditional network slicing control and service management and combined with the distributed computing and cache capabilities provided by the F-RAN architecture, this paper proposes a new type of network slice process. Different from traditional network slicing procedure, such procedure can respond to delay-sensitive services and reduce the fronthaul network load, and it could be used as a solution for network slicing in the F-RAN architecture.

Keywords: Network Slice, Fog Computing, F-RAN

1. Introduction
Facing the plentiful application scenarios and business requirements in 5G era, mobile communication networks requires new network technologies as well as network architecture. In addition to massive MIMO, beamforming and other physical layer technologies, many has study on advanced radio access network architecture. In addition, network slicing, as an important enabling technology for 5G networks, aims to provide users with end-to-end, customized, and isolated network services, which can improve the utilization of network resources and improve the flexibility and security of the network. In the future network architecture, all objects with certain access and control rights to the infrastructure can be regarded as a tenant. And network slicing enables all tenants to provide services to UEs on shared virtual network resources. Therefore, network slicing has the following benefits:

- Improve the service quality of user equipment.
- Improve the utilization of network resources.
- Reduce the complexity of network management and control for operators.
- Improve the revenue model of mobile communication network.

SliceNet is a project of 5G PPP which focus on managing controlling and orchestrating of 5G network slicing[1]. In addition, many standardization organizations have started research on network slicing. Network Functions and core network slice of the 5G core network has been defined in the 3GPP specification [2]. Network slicing has also caused a wave of research in academia. Rafael Montero et al. proposed an end-to-end network slice resource orchestration architecture [3]. Some scholars have applied deep reinforcement learning to network slicing resource management [4]. The
current research status of network slicing is summarized in [5], which also proposed a service management and network slicing control procedure. In the traditional network slicing procedure, the creation of a slice requires multiple mappings between multiple control layers, which is difficult to deal with delay-sensitive services. And it will generate amount of computing expenses in massive access scenarios, which will put tremendous pressure to the server.

Benefiting from SDN and NFV technologies, China Mobile proposed the C-RAN architecture as an improvement of the traditional access network architecture in 2013 [6]. The C-RAN architecture has multiple benefits, such as high spectrum utilization and lower energy consumption. However, the coupling of data and control increases the complexity of service control and deployment in the dense deployment scenario of base stations. To solve this problem, H-CRAN was proposed which combines macro-micro heterogeneous networking and C-RAN architecture [7], and the SDN-based FlexRAN was also proposed [8]. Yet both H-CRAN and FlexRAN architectures face the problem that network performance is limited by non-ideal fronthaul networks when densely deployed base stations. For this reason, people have proposed F-RAN based on the fog computing platform as a solution for 5G wireless access network [9]. The excellent characteristics of F-RAN can perfectly meet the requirements of 3GPP for 5G access networks [2]. Therefore, it is of great significance to propose a new network slicing procedure in F-RAN.

In this article, we first analyzed the existing network slicing process, then introduced the F-RAN architecture and its improvements, and finally we proposed a bottom-up network slicing process under the F-RAN architecture.

2. Network slicing and its procedure

In the next generation mobile communication networks, the realization of a commercial service consists of four departments, namely user equipment, infrastructure provider, network operator and application service provider. The infrastructure provider abstracts the hardware facilities and provides them to the network operator. Afterwards the network operators can perform network slicing for UEs in different application scenarios on shared network infrastructure based on SLAs and business requirements. A network slicing procedure based on 5GPPP network architecture [10] was proposed in [5]. In this procedure, the implementation of the network slice is divided into service management layer and network slice control layer. The specific procedure is as follows:

1) After receiving the network slice request, the network slice controller layer abstracts the resources and provides them to the service management layer.
2) The service manager performs abstraction, negotiation, and admission control on vertical industry applications (transportation, agriculture, and automated factory etc.) or third-party applications.
3) The service manager finds an appropriate slice template and creates a business class based on the SLA or based on the settings of the application service provider.
4) The service management layer feeds back the configuration information of this business class (VNFs, PNFs, additional services, security mechanisms, etc.) to the network slice controller.
5) The network slice controller then instantiates the slice resource according to the received configuration information.
6) The network slice controller performs slice selection and user attachment, and completes the network slice creation.

In addition, the network slice controller also needs to perform multiple slice interconnections, system status monitoring and analysis, slice reconfiguration, and other functions to optimize the overall network performance. However, network slicing under this architecture also has its own limitations. First, the service management and slicing control are using different controller, thus multiple mappings between controllers are required, which will undoubtedly bring huge computing overhead in the mass access scenario. Second, the performance of the centralized controller is limited by the transmission performance of the fronthaul network. Finally, after the slice request reaches the service management layer, it starts to negotiate with the application service provider to create a
business class, therefore, this procedure is difficult to meet the service requirements of ultra-low latency.

3. Proposed network slicing procedure in F-RAN

3.1. Fog computing & F-RAN
Cisco proposed fog computing extends computing and caching capabilities to the edge of the network, enabling new applications and services. Fog computing has characteristics such as mobility, location awareness, and heterogeneity[11]. Compare to cloud computing, fog computing nodes are closer to user equipment, and it can provide lower latency, thereby they are very suitable for 5G access networks. M Peng et al first proposed the fog-computing-based mobile access network architecture F-RAN as a 5G RAN solution [9].

All user equipment with caching and computing capabilities in F-RAN can be used as fog-computing-based user equipment (F-UE). All F-UEs, fog-computing-based access points (F-APs) and cloud computing server located in BBU pool can perform collaboration radio signal processing (CRSP) and cooperative radio resource management (CRRM). That means, it can be orchestrated by a centralized controller to obtain a better user experience and higher resource utilization. The architecture of F-RAN is shown in Figure 1.

![Figure 1. 5G F-RAN architecture.](image)

This architecture can meet the requirements of 3GPP for future communication network architectures [8]. First, F-RAN can separate the user plane and control plane. Second, network functions in F-RAN can interact with each other. Third, F-RAN can minimize the dependency of the access network and the core network, which means F-UEs can use a public interface to access the core network. Fourth, this architecture can deploy some user plane functions near the access network, which supports concurrent access to local and centralized services. In addition, F-RAN also supports
capabilities exposure. Therefore, the F-RAN architecture can be used as a good solution for 5G access networks.

3.2. Proposed Bottom-up Network Slicing Procedure in F-RAN

In order to achieve the best transmission performance, F-UEs need to choose between four transmission modes. An adaptive transmission mode selection algorithm based on the distance between F-UEs and whether it supports D2D communication mode was proposed in [12]. In such algorithm, if the F-UEs are in the D2D communication mode, the F-UEs acting as cluster heads need to buffer the data of all the surrounding UEs. Therefore, the communication quality and caching capabilities of the F-UEs have strict requirements. In this architecture, F-UEs and F-APs can provide additional computing capabilities to deploy some machine learning and AI algorithms to optimize the overall network performance. For instance, by deploying content awareness at F-Aps and the user equipment attaches its own status information on the slice request, so that F-APs can classify the service locally after receiving the slice request from the user equipment. Then the F-AP encodes the F-UE's location information \( LI \), service category \( SC \), channel information \( CF \) and other related information into \( X = \{(LI), (SC)\} \), and perform K-means clustering for all users on \( X \). By selecting the appropriate \( K \) to make sure the number of categories of access users is as small as possible, and users in the same category are as close as possible in both spatial location and service category. Then, the cluster head is picked from all the receiving user clusters, and all receiving users with similar service types are replaced by the cluster head, thereby reducing the fronthaul network load and replacing the hardware requirements of the cluster head.

The centralized control cloud server at the BBU pool has a global view, including the remaining cache, remaining computing capabilities, and power status of all F-APs in its coverage area, as well as the location status and channel status of the corresponding access user equipment. Undoubtedly, the powerful computing ability of the centralized control cloud can be used to deploy reinforcement learning models to quickly complete the adaptive mapping of slice instances to physical resources when new slice instances are formed. And the control cloud server can cache hotspot services and delay-sensitive business templates in the coverage area locally or closer to the user device in advance. When a new slice request arrives, it can quickly complete the creation of slice instances, thereby further reducing delay. Dependent on the F-RAN architecture with powerful caching and computing functions, a new type of bottom-up network slicing procedure can be proposed. The specific slicing process is shown in Figure 2. The specific slicing steps are:

1) When the UE has a new service request, it attaches its own status information to the slice request and sends it to the F-APs.

2) The local server at the F-APs obtains the channel state of the UE through the reference signal, and then obtains the service category, then all access users are K-Means clustered on \( X = \{(LI), (SC)\} \), and select the appropriate cluster head according to user status, combined into a small-scale slice request message and sent to a centralized cloud server located in the BBU pool.

3) F-AP abstracts the resource information and sends it to the centralized control cloud server.

4) For popular and delay-sensitive services, the centralized cloud server refers to the local slice template to create slice instances. For other services, the centralized control cloud server negotiates with the application service provider to obtain its slice instances.

5) The centralized control cloud server schedules the corresponding resources for the created slice instance according to the resource configuration information and sends it to the base station corresponding to the user equipment.

6) After the configuration of the relevant resources of the slice is completed, the user attachment is performed, and the creation of the slice is completed up to this point.

It can be obtained that the procedure has the following advantages. First, through K-Means clustering in combined with service categories at F-APs, the traditional distance-based clustering
algorithm is improved, reducing the hardware requirements for cluster heads and reducing the load on the fronthaul link. Secondly, the centralized control cloud server under the F-RAN architecture can replace the network slicing controller and service manager in the traditional network slicing process, eliminating the complicated resource mapping process between the two and reducing the server's computing pressure. Finally, the use of the cache computing capabilities of fog computing nodes can achieve the agile deployment of delay-sensitive service slicing, as well as reduce the number of transmissions of popular services in the network to reduce the network load.

Figure 2. F-RAN bottom-up network slicing process.
4. Conclusion

Facing the complex and diverse application scenarios and business requirements of 5G era, this paper proposes a bottom-up network slicing process in the new network architecture F-RAN. This new procedure can reduce latency, and reduce network load with lower server computing costs.

In the future, with the additional computing power provided by F-RAN, some AI algorithms can be used to solve problems such as virtual network embedding, adaptive network resource scheduling and network reconfiguration, and promote the intelligent development of mobile communication networks.

References

[1] Q. Wang et al., “SliceNet: End-to-End Cognitive Network Slicing and Slice Management Framework in Virtualised Multi-Domain, Multi-Tenant 5G Networks,” IEEE Int. Symp. Broadband Multimed. Syst. Broadcast. BMSB, vol. 2018-June, 2018.

[2] System Architecture for the 5G System,” “3GPP TS 23.501” Release 16, 2019.

[3] R. Montero, F. Agraz, A. Pages, and S. Spadaro, “End-to-End 5G Service Deployment and Orchestration in Optical Networks with QoE Guarantees,” Int. Conf. Transparent Opt. Networks, vol. 2018-July, pp. 1–4, 2018.

[4] R. Li et al., “Deep Reinforcement Learning for Resource Management in Network Slicing,” IEEE Access, vol. 6, pp. 74429–74441, 2018.

[5] I. Afolabi, T. Taleb, K. Samdanis, A. Ksentini, and H. Flinck, “Network slicing and softwarization: A survey on principles, enabling technologies, and solutions,” IEEE Commun. Surv. Tutorials, vol. 20, no. 3, pp. 2429–2453, 2018.

[6] C. Mobile, “C-RAN: the road towards green RAN,” White Pap. ver 3.0, vol. 0, pp. 15–16, 2015.

[7] M. Peng, Y. Li, J. Jiang, J. Li, and C. Wang, “Heterogeneous cloud radio access networks: A new perspective for enhancing spectral and energy efficiencies,” IEEE Wirel. Commun., vol. 21, no. 6, pp. 126–135, 2014.

[8] 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,” Conex. 2016 - Proc. 12th Int. Conf. Emerg. Netw. Exp. Technol., pp. 427–441, 2016.

[9] M. Peng, S. Yan, K. Zhang, and C. Wang, “Fog-computing-based radio access networks: Issues and challenges,” IEEE Netw., vol. 30, no. 4, pp. 46–53, 2016.

[10] 5G-PPP Architecture Working Group, “5G-PPP White Paper on 5G Architecture,” no. Dec 2017, p. 140, 2017.

[11] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the internet of things,” MCC ’12 - Proc. 1st ACM Mob. Cloud Comput. Work., pp. 13–15, 2012.

[12] S. Yan, M. Peng, and W. Wang, “User access mode selection in fog computing based radio access networks,” 2016 IEEE Int. Conf. Commun. ICC 2016, no. February, 2016.