Review Article

5G Network Slicing to Deliver Improved User Experience

Mudit Sood

Cellular Verification Architect, Apple.

Corresponding Author : muditsood@gmail.com

Received: 18 July 2023

Revised: 30 August 2023

Accepted: 13 September 2023

Published: 30 September 2023

Abstract - 5G technologies have further advanced cellular to the frontier of wireless communications. 5G has brought several compelling use cases to life, like eMBB, uRLLC and massive IOT. Network Slicing will allow creative use of these features over the same bandwidth. Carriers will be able to slice the huge bandwidths that 5G allows into manageable distinct parts and associate appropriate QoS parameters to each part to suit the type of traffic that they will carry, thus creating a Virtualized network. With increased flexibility comes complexity, and many different Communication Service Provider companies have standardized and launched services to make the end-to-end realization of the use cases possible. This paper provides an overview of the technology behind Network Slicing, API Framework, Monetization and some other considerations to bring the power of network slicing to the customers and current research gaps. It also shows how Applications can utilize network Slicing to enhance the user experience of their customers.

Keywords - 5G Network Slicing, eMBB, URLLC, mMTC, Massive IOT, End to end user Experience, Network Function virtualization.

1. Introduction

ITU defined the roadmap and requirements to reach "IMT-2020". [3] 3rd Generation Partnership Project (3GPP) is the organisation that provides standards governing 5G and all its features and met and exceeded the IMT-2020 requirements with 5G technology. 5G has built on top of previous generations and brought industry-leading research and findings to life. One of the key features of 5G, Network slicing, will create a virtual network within the existing infrastructure of a carrier, allowing the UE to request slices of different service type (ST) and Service Differentiator (SD), which will guarantee specific Service Level Agreement (SLA). Communication Service will interface with applications running on the device and further extend the flexibility so that applications can meet a high level of user experience. It also opens the field to Mobile Virtual Network Operators (MVNO), thus bringing innovation to solve issues related to rolling out complex 5G features.

2.5G Use Cases

IMT-2020 requirements introduced 3 use cases which can layout the flexibility offered by 5G. Many other use cases have been defined and are deployed under 5G, which include some aspects from below 3 in addition to other details. For the purpose of this paper, these 3 will provide the reader with a good view of features that can then be leveraged in a given network slice. 3GPP also provides standards for other access network types (like WLAN) to realise Network slicing. UE can request a slice which provides benefits of one of these use cases end to end. Which slice to choose will depend on the application using the connection. The figure below shows use cases on vertices of a triangle, highlighting applications inside the triangle and closer to the relevant use case vertex.



Fig. 1 5G use cases and some applications Source: ETRI graphic from ITU-R IMT 2020 requirements

2.1. Enhanced Mobile Broadband (eMBB)

eMBB will provide gigabit speeds and more capacity while the device is on the move. Think of streaming highdefinition movies and AR and VR applications on mobile devices while travelling in a high-speed train along with everyone else on the train. This is achieved by bigger bandwidths, beamforming and higher spectral efficiency in 5G. eMBB is the most prevalent and first use-case to be implemented by carriers.

2.2. Ultra-Reliable Low Latency Communication (URLLC)

This use case covers low lag and highly reliable connection, enabling mission-critical applications like vehicle-to-vehicle communication for autonomous driving and remote surgery. This is enabled by making a flexible radio frame structure, allowing features like self-contained slots where transmission and acknowledgement over the air can be packed in less than 1 ms. With advanced wireless technology and improved receiver design, 3GPP has placed stricter UE processing time requirements, creating frequent occasions for control and data over the air. End-to-end latency from the application to the server and back is less than 5 ms. Improved coding and gains from beamforming allow ultra-high reliability.

2.3 Massive IOT

Massive Internet of Things (IOT) or massive Machine Type Communication (mMTC) will allow more than a million devices in one square kilometre to be connected. These devices will transmit low data rates and should consume low power to be able to operate for up to 10 years. Some applications are automated factory floor operations, freight tracking, and electric meters. Special provisions by 3GPP standards like Power Saving Mode achieve this.

3. Network Slicing

Network Slicing is 3GPP defined architecture that will allow Mobile Virtual Network Operators (MVNO) to define, deploy, operate and monetize a virtual network that can meet their customers' specific needs and the devices being served in that network. The device and the network serving it must implement all the requirements laid out in 3GPP and can utilize the parameters to enhance the user experience.

An MVNO can design network slices based on their customers' requirements or their products' functions. They should use S-NSSAI according to the services offered. The identity convention is flexible to allow for many different use cases. They may have use where the same services are offered via different S-NSSAIs, each to a group of UEs belonging to different customers or applications. In that case, the network slices will have the same N-SSAI and respective service type and will be differentiated by different service identifiers. A UE can have multiple network slices active simultaneously, serving different applications with specific requirements for the connection. These network slices could be with the same or different MVNOs and could be using a 5G access network or even a non-3GPP access like WiFi. 3GPP architecture is flexible to allow all such and many more use cases. [4]

4. 3GPP Parameters and Interfaces

Parameters must be negotiated and communicated for the various nodes involved in a network slice end-to-end to work seamlessly. These parameters control the characteristics of the connection and resources needed at each node to conform to the QoS requirements, or SLA agreed upon by the provider. Below are parameter definitions from 3GPP standards. [4]

4.1. NSSAI

Network Slice Selection Assistance Information (NSSAI) specifies information about Network Slice. It consists of 8 bits of Slice/Service type (SST) and 24 bits of Slice Differentiator (SD). Table 1 shows values currently specified for SST. UE requests an S-NSSAI at the time of starting the connection. The network responds with a set of allowed N-SSAI for the UE and connection. At the time of PDU session establishment, selected S-NSSAI is used from the intersection of requested and allowed S-NSSAI. SD part is a 3-octet string to differentiate amongst multiple Network Slices of the same Slice/Service type.

4.2. Network Slice Instance Identifier (NSI ID)

A Network Slice Instance Identifier (NSI ID) uniquely identifies a Network Slice Instance (NSI) within a Public Land Mobile Network (PLMN) or Standalone Non-Public Network (SNPN) when multiple NSIs of the same Network Slice are deployed and there is the need to differentiate between them in the 5GC. An NSI may be associated with one or more S-NSSAIs and an S-NSSAI may be associated with one or more NSIs. [4]

Slice/Service type	SST value	Characteristics
eMBB	1	Slice suitable for the handling of 5G enhanced Mobile Broadband.
URLLC	2	Slice suitable for the handling of ultra-reliable, low-latency communications.
МІоТ	3	Slice suitable for the handling of massive IoT.
V2X	4	Slice suitable for the handling of V2X services.
НМТС	5	Slice suitable for the handling of High-Performance Machine-Type Communications.
HDLLC	6	Slice suitable for the handling of High Data rate and Low Latency Communications.

Table 1. Standardized SST and its characteristics [4]

4.3 Interfaces involved in Network Slicing

Network Slice Instance has a scope with a PLMN or SNPN and has associated Network Functions to perform control plane and user plane operations. Control plane setups up all the nodes in the network slice based on the configured S-NSSAI. The user plane provides resources for the data to flow in both directions from an application running on UE to the Application Server in the core network and the other way. The network slice also needs access to the network to reach UE. This can be a 5G Radio Access Network or non-3GPP. 3GPP has defined interfaces towards such non-3GPP entities, like Non-3GPP Inter Working Function (N3IWF) or Trusted Non-3GPP Gateway Function (TNGF) or Trusted WLAN Interworking Function (TWIF) functions to the trusted WLAN in the case of support of Non-5G-Capable over WLAN (N5CW) devices. There is even the Wireline Access Gateway Function (W-AGF) to the Wireline Access Network.

5. 3GPP and Mobile OS API Interfaces

Application Programming Interface (API) is instrumental in implementing flow for a complex system. Independent parties can implement components of the system. However, for each of these components to work seamlessly, they should be able to communicate with each other and serve the functions implemented by the independent logic. 3GPP has specified a Common API Framework (CAPIF) for this purpose so that companies involved in the business of providing solutions for features like Network Slicing are consistent in their approach and are interoperable with each other. The CAPIF framework includes several functions critical for designing, launching and operating services for companies. It also has monitoring functions allowing analytics for usage, planning, sales and security-related aspects. [13]

CAPIF includes common aspects applicable to any northbound service APIs from 3GPP. Northbound means APIs that face towards business logic or applications that are customer-facing. These are the APIs which the end application running on the mobile device can take advantage of to make the user experience better. For example, an online gaming application with users playing high-definition games in real-time will have strict latency and reliability requirements for the network connection. If such an Application is aware of the CAPIF services and has required permissions, it can invoke API with the right parameters that meet latency and reliability limits. CAPIF provider will, in turn, call the service API provider, which will make the UE send a connection request with S-NSSAI service type as 2 for URLLC. The figure below shows high-level parts involved in Resource Owner Aware Northbound API Access (RNAA) and the relationships between them. CAPIF and API providers will normally be implemented together by one company but have a standard interface to allow flexibility.

5.1. Definitions for the entities

A 3rd party application provider typically provides the API invoker with a service agreement with the PLMN operator. The API invoker may reside within the same trust domain as the PLMN operator network. [13] The API invoker may be either an application on a server or an application on a UE. CAPIF provider implements the 3GPP specified interface and will use the published yaml to communicate with the API invoker. It acts as an abstraction to the actual API provider so that the service provided by the API can be invoked in a standardized way and the service agreement can be negotiated. For example, on an Android device, the CAPIF provider interface will be implemented by the Android Open Source Project (AOSP) telephony framework. Applications with required permissions can leverage such APIs.



Fig. 2 Business Relationships for Resource Owner Aware Northbound API Access (RNAA) [13]

Carriers (like AT&T Verizon) can differentiate their applications by allowing exclusive access to such APIs. They can monetize it by selling premium subscriptions to their customers, or they can license the use of such APIs to 3rd parties. In this example, the carrier would implement the rest of the parts (API provider and resource owner).

API provider can be implemented by another company that might specialize in specific use cases like freight tracking. Such a company may not own any core network infrastructure and can utilize a *Resource Owner* to bundle together and sell the services through a commercial CAPIF API. There will be a Resource Access Arrangement, another standard interface that would allow authentication, authorization and charging to allow the use of core network infrastructure where the API logic will actually run. Such modular architecture allows for greater flexibility and opens up opportunities for start-ups and smaller companies focusing on specific areas to collaborate and deliver complete solutions to applications. Since there are many interfaces and possibilities, network slice management is crucial to ensure smooth operation throughout the life cycle of the network slice. A lot of research is happening to utilize Machine Learning (ML) to manage and orchestrate all the transactions happening within a network slice. Doing this with human intervention will be tedious and error-prone and become a bottleneck as the number of deployments increases.

5.2. Network Slicing APIs in AOSP

Android Open Source Project (AOSP) is Google's version of the Android operating system for mobile devices optimized for touchscreen operation. Many flagship devices manufactured by OEMs ship with AOSP as the base operating system. Like several different frameworks inside AOSP, Telephony Framework provides telephony functionalities like voice calls, data calls, SMS, etc, to mobile devices. Under connectivity, there is the framework for Network Slicing.



Fig. 3 5G network slicing architecture in AOSP [14]

Many Original Equipment Manufacturers, including world-leading companies like Samsung, have built mobile devices on top of AOSP. With such a large user base to deploy services, companies can find customers for their business use cases. A couple of ways that the traffic for such applications on work devices can be routed to the Network Slice are mentioned below. [14]

5.1.1 Fully-managed devices

These are devices the company provides to their employees and manages by their IT. These can be mobile phones or devices with WiFi access. They will be programmed to connect to the virtual network, and data from the work apps will get routed over network slice, getting the configured QoS and policing to comply with the enterprise rules. [14] User Equipment Route Selection Policy (URSP) is the method AOSP and Google employ to allow enterprises and carriers to control what applications can use the network slices and how.

5.1.2 Enterprise business app slicing for devices with work profiles

For enterprises using the work profile solution, Android 12 allows devices to route the traffic from all apps in the work profile to an enterprise network slice. Enterprises can enable this capability through a Device Policy Controller

(DPC). The work profile solution provides an automatic level of authentication and access control that enterprises require to ensure that only traffic from enterprise apps in the work profile is routed to the enterprise network slice. Apps in the work profile do not need to be modified to request the enterprise network slice explicitly. With this method, employees can use a single device to personalise and work with regular apps using the generic connectivity from the device and work applications routed through the enterprisecontrolled virtual network. [14]

Network Slicing allows carriers to gain control over how applications use their network resources and monetize their use. Several popular applications run over the top of cellular networks and get the benefit of ubiquitous, reliable cellular connectivity and charge the end user while the data goes over carriers' unlimited data plan. However, carriers can build a case of a quarantined level of service, differentiate their native applications and license them to 3rd party applications. This is again possible due to layered architecture.

5.1.3. URSP Rules Allow Carriers to Control and Monetize Network Slicing

Carriers can configure URSP rules for different slice categories, including enterprise, CBS, low latency, and high bandwidth traffic. When configuring URSP rules for different slice categories, carriers must include OSId for Android, a version 5 Universally Unique Identifier (UUID) generated with the namespace ISO OID and the name "Android". AppID will identify the App type like ENTERPRISE, CBS, etc. URSP will have OSId+AppID as a field. [14]

Table 2. Example of a URSP rule specified by the carrier. [14]		
URSP rule #1 (enterprise1)		

UKSF Tule #1 (enterprise1)			
Precedence	1 (0x01)		
Traffic descriptor #1			
OS Id + OS App Id type	0x97A498E3FC925C9489		
	860333D06E4E470A454E		
	5445525052495345		
Route selection descriptor #1			
Precedence	1 (0x01)		
Component #1: S-NSSAI	SST:XX SD: YYYYYY		
Component #2: DNN	enterprise		
Route selection descriptor #2			
Precedence	2 (0x02)		
Component #1: DNN	enterprise		

With standardization and provisions like URSP rules, Network Slicing adoption will get a boost. This framework allows informed applications to take advantage and really enhance the user experience for its customers. It makes many use cases possible due to the QoS and QoE agreements it can support.



Fig. 4 Communication services running over network slices. [8]

4. End-to-end QoS and QoE

3GPP has several provisions in air interface design and enhancements in the core network to support Network Slicing. In order to meet these requirements end-to-end, each system in the flow has to prioritize the traffic for that connection such that the end user gets the promised service. Each layer provides Quality of Service (QoS) from the input to the output. However, to guarantee Quality of Experience (QoE) for end users, all the layers and modes must work harmoniously to deliver the service. Another challenge is translating the QoS parameters like latency data rate to QoE parameters like responsive, intuitive and low failure rate is challenging. ITU and 3GPP provide guidance in this aspect; however, companies in this business have to come up with their own solutions to gain a competitive advantage.

Communication service is a contract with the end user, which can be an application for delivering traffic within the limits of the data rate, reliability, latency, and connection density. With all the interfaces standardized by 3GPP, different entities can deliver products and services to complete the flow. The communication service provider implements the communication service. The core network has to be flexible enough and should have the capacity to serve traffic going over several communication services, each with different characteristics. Network Function Virtualization achieves this. Functions needed to perform a specific task are broken into logical Network Functions (NF), and interfaces are defined for it to communicate with the rest of the functions and the system. With such component functions defined, several companies can work together to deliver the solution as a whole. Network Slice Selection Function (NSSF) to select a set of Network Slice instances serving the UE. Network Slice Specific Authentication and Authorization Function (NSSAAF) to support authentication and authorization with a AAA server. Support for access to a Standalone Non-Public Network (SNPN). Network Slice Admission Control Function (NSACF) monitors and controls the number of registered UEs per network slice and/or the number of PDU Sessions per network slice for the network slices subject to Network Slice Admission Control (NSAC). [4]

4.1. Network Slice as a Service (NSaaS)

NSaaS is a service offered by companies who want to leverage the 5G network features customized for their application. They do not want to invest in infrastructure or know how to configure and manage a Network slice. Communication Service Provider can provide an end-to-end solution or expose some management interfaces to allow more control over the aspects of Network Slice. [8]

4.2. Network Slices as NOP Internals

In the "network slices as Network Operator (NOP) internals" model, network slices are not part of the NOP service offering and are not visible to its customers.

However, to provide support to communication services, the NOP may decide to deploy network slices, e.g., for internal network optimization purposes. This model allows the Communication Services Customer (CSC) to use the network as the end user or optionally allows the CSC to monitor the service status. [8]

Quality of Experience (QoE) measures customer satisfaction with the communication service. Great QoE will lead to a great overall user experience. QoE is an overarching measure requiring each network component to meet the QoS agreement. At the same time, there may not be a one-to-one and linear relationship between the parameters of QoE and the parameters of QoS. One example of this is voice calls on cellular networks. In this case, QoS parameters will be the RTP jitter packet loss. QoE is a subjective measure, and one such measure is the Mean Opinion Score (MOS). Even when these QoS parameters are within limit, good QoE will require proper choice of codecs corresponding to the. Content is being transferred and end-to-end provisioning, so there is no lag or call drop.

In order to maintain good QoE and predict when it might degrade, it is important to understand the relationship between the two. This can be achieved by building a data model of the end-to-end services. Carriers or Application companies offering such end-to-end services can conduct user surveys, collect field diagnostics, and then overlay them with the systems' conditions and indicators of health and measure the QoS parameters. This can help predict any service degrades that might lead to an outage or bad user experience and take pre-emptive steps to avoid them.

4.3. Network Slice Charging

3GPP has standardized entities, roles and interfaces to allow for different charging mechanisms for usage of services via Network slice. [15] This will allow several players to monetize their service offerings to entities around them, like Network operators (NOP), Communication Service Providers (CSP), and CAPIF API providers. Users will also benefit from such flexible charging mechanisms. They could upgrade their service for a better User Experience or shop from competitive offerings when buying new subscriptions or international roaming.

5. Full Stack Testing

The network slicing concept runs through many layers in the 3GPP air interface, core network and application servers. In order to ensure they all are inter-compatible, these should be tested in full stack setup. The below figure shows the setup for testing implementation on UE. The arrow showing New Radio (NR) represents the air interface for the 5G network. This will have the real device launched on the UE side and the network simulator on the other side. All the air interface features relevant to the Network slicing use case being tested can be applied and verified in this setup. Above the physical layer is the network and TCP/IP, which provides the application with reliable and effective transport service. Since the scope here is testing the network slice implementation, the application on the UE side can be a real or a simulated one with an interface to interact with the commendation service APIs.



Fig. 5 Full Stack Testing for Network Slicing [11]

On the network side, the system simulator in the green box will have the test equipment to simulate the NR interface. This test equipment acts as an access network, and UE communicates with it as it would with a real network in the field. So, all the UE side radio components must exercise the field configuration. Test equipment connects to Application server simulators over TCP/IP, the data plane via which the application on the UE side will be served data or services. The other interface is to the Main Controller, where test config to control the network side behaviour can be applied to the system.

Full stack testing ensures end-to-end field performance of products and services built over Network slicing. There are so many interfaces, and many companies may be involved in providing solutions that form the Network slice together. Full stack testing is a way to make sure that they will work when put together in the field.

6. Research Gaps

Network Slicing has solved the problem that applications cannot use a one-size-fits-all all-network, be responsive, and deliver a great user experience. However, in doing so, it has exposed some vulnerabilities. Security gaps include un-authorized access, data breaches and denial of service attacks. Since Network Function Virtualization (NFV) is used extensively for Network Slicing, it adds to the inherent challenges that NFV has. [10] Orchestrating solutions across various vendors presents inter-operability issues. Managing so many nodes across the life cycle of a Network Slice from inception, commissioning, operation, and de-commissioning has become very tedious and prone to errors. AI/ML algorithms are being researched and deployed to solve the challenges of orchestrating. However, the completeness and effectiveness of such algorithms are limited due to the limited data available to train them.

7. Conclusion

Through the facts and information in this paper, it can be concluded that network slicing in 5G can deliver an excellent user experience as promised by the standards. There are enough implementations of the component functions to make it a reality. Standards defined for non-3GPP access networks can extend the application user base to users with WLAN or Wired LAN access, opening the field to diverse uses and adding more value for the end user. This paper introduced high-level 5G use cases and defined network slicing along with 3GPP parameters associated with it. Next, it presented what it takes to implement Network slice end-to-end, including architecture for API interfaces and examples from AOSP. This paper also provides considerations for designing and monetizing products and services built over Network Slicing. It explored some of the gaps in current research.

References

- [1] "ETSI TS 138 401. 5G; NG-RAN; Architecture Description (3GPP TS 38.401 Version 16.3. 0 Release 16)," Technical Specification, pp. 1-80, 2020. [Google Scholar] [Publisher Link]
- [2] Numbering, Addressing and Identification, Technical Specification 23.003, Version 18.3.0, 3GPP, 2023. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=729
- [3] Eiman Mohyeldin, "Minimum Technical Performance Requirements for IMT-2020 Radio Interface(s)," *ITU-R Workshop on IMT-2020 Terrestrial Radio Interfaces*, pp. 1-12, 2020. [Google Scholar] [Publisher Link]
- [4] System Architecture for the 5G System (5GS), Technical specification 23.501, Version 18.3.0, 3GPP, 2023. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144
- [5] Rao Yallapragada, and Jing Jiang, "New Services and Applications with 5G Ultra Reliable Low Latency Communication," 5G *Americas*, pp. 1-60, 2018. [Google Scholar] [Publisher Link]
- [6] Saoussen Chaabnia, and Aref Meddeb, "Slicing Aware QoS/QoE in Software Defined Smart Home Network," *NOMS 2018 2018 IEEE/IFIP Network Operations and Management Symposium*, pp. 1-5, 2018. [CrossRef] [Google Scholar] [Publisher Link]
- [7] Faqir Zarrar Yousaf et al., "Network Slicing with Flexible Mobility and QoS/QoE Support for 5G Networks," 2017 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1195-1201, 2017. [CrossRef] [Google Scholar] [Publisher Link]
- [8] Mirko Cano Soveri et al., 5G Network Slice Management, 3GPP, 2023. [Online]. Available: https://www.3gpp.org/technologies/slicemanagement
- [9] 5G Network Slicing: Security Considerations for Design, Deployment, and Maintenance, Media Defense. [Online]. Available: https://media.defense.gov/2023/Jul/17/2003260829/-1/-1/0/ESF%205G%20NETWORK%20SLICING-SECURITY%20CONSIDERATIONS%20FOR%20DESIGN,%20DEPLOYMENT,%20AND%20MAINTENANCE_FINAL.PDF
- [10] Shilpa Bhairanatti, and S. Mohan Kumar, "Evolution of 6G Era: A Brief Survey of Massive MIMO, mm-Wave, NOMA-Based 5G and 6G Communication Protocols, Role of Deep Learning and Inherent Challenges," SSRG International Journal of Electrical and Electronics Engineering, vol. 10, no. 1, pp. 24-40, 2023. [CrossRef] [Publisher Link]
- [11] Technical Report 38.918, Study on 5G NR User Equipment (UE) Full Stack Testing for Network Slicing, Version 18.1.0, 3GPP, 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3891
- [12] "Common API Framework for 3GPP Northbound APIs, 3GPP TS 23.222 version 15.3.0," 3GPP, pp. 1-87, 2018. [Google Scholar][Publisher Link]
- [13] Common API Framework for 3GPP Northbound APIs, Technical specification 23.222, Version 18.2.0, 3GPP, 2023. [Online]. Available: https://portal.3gpp.org/desktopmodules/SpecificationSpecificationDetails.aspx?specificationId=3337
- [14] AOSP Documentation: 5G Network Slicing, Android Source, 2023. [Online]. Available: https://source.android.com/docs/core/connect/5g-slicing
- [15] 3GPP TR 32.847, Study on Charging Aspects for Network Slicing Phase 2, 3GPP. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3842