ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΜΗΧΑΝΙΚΗ ΥΠΟΛΟΓΙΣΤΩΝ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ ΚΑΙ ΔΙΚΤΥΩΝ"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

A Comprehensive Survey on 5G Adaptive Network Slicing: Key Technological Principles, Deployment and Operation Challenges

Ευάγγελος Γ. Σιατήρας

Επιβλέπων: **Δημήτριος Τσόλκας**, Διδάσκων

ΑΘΗΝΑ

ΙΟΥΝΙΟΣ 2021

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

SCHOOL OF SCIENCES
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS

PROGRAM OF POSTGRADUATE STUDIES IN COMPUTER TELECOMMUNICATIONS AND NETWORK ENGINEERING

MSc THESIS

A Comprehensive Survey on 5G Adaptive Network Slicing: Key Technological Principles, Deployment and Operation Challenges

Evangelos G. Siatiras

Supervisor**:** **Dimitrios Tsolkas**, Adjunct Teaching Stuff

ATHENS

JUNE 2021

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

A Comprehensive Survey on 5G Adaptive Network Slicing: Key Technological Principles, Deployment and Operation Challenges

Ευάγγελος Γ. Σιατήρας
Α.Μ.: EN2190001

ΕΠΙΒΛΕΠΩΝ:	**Δημήτριος Τσόλκας**, Διδάσκων

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:	**Λάζαρος Μεράκος**, Καθηγητής **Νικόλαος Πασσάς**, Μέλος Ε.ΔΙ.Π. **Δημήτριος Τσόλκας**, Διδάσκων

ΙΟΥΝΙΟΣ 2021

ΠΕΡΙΛΗΨΗ

Ο κλάδος των τηλεπικοινωνιών μεταβαίνει από ένα "οριζόντιο" μοντέλο παροχής υπηρεσιών στο οποίο οι υπηρεσίες καθορίζονται ανεξάρτητα από τις ανάγκες των πελατών τους και κατευθύνεται προς ένα "κάθετο" μοντέλο παροχής υπηρεσιών στο οποίο οι υπηρεσίες που προσφέρονται διαμορφώνονται ανάλογα με τις ανάγκες των καταναλωτών. Για να καταστεί δυνατή αυτή η μετάβαση, απαιτείται μια ισχυρή υποδομή 5G από άκρο σε άκρο, με τεχνολογίες ικανές να εξυπηρετούν τις διάφορες περιπτώσεις χρήσης κάθε βιομηχανίας. Ένα κρίσιμο στοιχείο αυτής της αρχιτεκτονικής είναι η ενσωμάτωση της έννοιας του τεμαχισμού του δικτύου (network slicing) πάνω από την υπάρχουσα υποδομή ενός Παρόχου για την παροχή υπηρεσιών που διαφέρουν μεταξύ τους σε μεγάλο βαθμό τόσο στα χαρακτηριστικά τους όσο και στην απόδοση, καθώς και ένα μηχανισμό διαχείρισης του τεμαχισμού του δικτύου ικανό να διαχειρίζεται ταυτόχρονα πολλές φέτες. Σε αυτή τη μελέτη, εμπεριέχεται μια ολοκληρωμένη συζήτηση για την έννοια και την αρχιτεκτονική του συστήματος του τεμαχισμού δικτύου από την οπτική γωνιά του Παρόχου αυτών των υπηρεσιών αναλύοντας κρίσιμες πτυχές διαχείρισης των Control, User, Management and Orchestration Planes. Θέλω να επισημάνω ότι ορισμένοι οργανισμοί καθορίζουν τα χαρακτηριστικά των δομών και διεπαφών που απαιτούνται για την ενσωμάτωση ενός μεγάλου αριθμού στοιχείων σε μια λειτουργική δομή που μπορεί να εφαρμοστεί στα συστήματα Παρόχου υπηρεσιών με βάση διάφορα επιχειρηματικά μοντέλα και περιπτώσεις χρήσης. Επίσης παρέχεται μια διεξοδική επισκόπηση ενός πρωτοτύπου ικανό να αναπτύξει τις φέτες που χρειάζονται οι διάφοροι πελάτες και να τους προσφέρει μια φιλική προς το χρήστη διεπαφή για να διαχειρίζονται τις φέτες του δικτύου που τους αντιστοιχεί. Τέλος παραθέτω κάποιες περαιτέρω ανοικτές ερευνητικές κατευθύνσεις και υπάρχουσες προκλήσεις, όπως ο διαχωρισμός της στοίβας πρωτοκόλλων της ραδιοεπαφής, ο διαχωρισμός των Over-the-Top εφαρμογών σε μικροϋπηρεσίες (microservices), την επίπτωση του τεμαχισμού του δικτύου στο επίπεδο μεταφοράς και τέλος θέματα λειτουργίας και διασφάλισης των παρεχόμενων υπηρεσιών με σκοπό την παρακίνηση για τον περαιτέρω εκσυγχρονισμό των μηχανισμών που χρησιμοποιούνται στην βιομηχανία και την επισήμανση ρεαλιστικών παρατηρήσεων γι’ αυτήν την αναδυόμενη τεχνολογία.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Τηλεπικοινωνίες

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: 5G SDN NFV Network slicing Cloud/edge computing Network Softwarization OpenSlice Network Management

ABSTRACT

The telecom industry is transitioning away from a "horizontal" service delivery model in which services are specified independently of their customers’ needs and heading towards a "vertical" service delivery model in which services are targeted to particular industry segments and verticals. To make this transition possible, a robust end-to-end 5G infrastructure is needed, with technologies able to serve the various vertical industries' use cases. A crucial element of this architecture is the integration of network slicing over a single infrastructure to provide incredibly diverse vertical services, as well as a network slicing management framework capable of managing several slices concurrently. In this study, comprehensive discussion is provided on concept and system architecture of network slicing with particular focus on its business aspect and how Control, User, Management and Orchestration Planes are addressed. A number of organizations are focusing on standardizing the design structures and interfaces needed for integrating a large number of components into a functional structure that can be applied within provider/operator systems based on a variety of business models and use cases. Furthermore, a thorough overview of a prototype is being presented which is capable of deploying the slices needed by the various vertical players and providing them with a user-friendly interface to handle their slice. The proposed product integrates additional modules to include improved management and orchestration and control capabilities, as well as a potential end-to-end testing platform, and expands on current standardization activities. Concluding with open research directions and existing challenges like Radio protocol stack disaggregation, Over-the-Top application disaggregation into microservices, the IP perspective of network slicing and operation and service assurance considerations with the purpose of motivating new advances and adding realistic remarks to this emerging technology.

SUBJECT AREA: Telecommunications

KEYWORDS: 5G SDN NFV Network slicing Cloud/edge computing Network Softwarization OpenSlice Network Management

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

This thesis provides a comprehensive review related to some aspects and considerations related to 5G network slicing and SDN and NFV mechanisms which are adopted inherently. Firstly, chapter 2.1, gives an overview regarding the 5G evolution through the main pillars standardized in Release 15, 16 and 17. Referring to some intelligence considerations in 5G architecture (chapter 2.2), one major improvement in the infrastructure of the Core network Architecture is the so-called Service Based Architecture (SBA). Furthermore, another interesting concept in the design of 5G Core is the orthogonality of functions. From a back-hole perspective (chapter 2.3) there is no significant difference comparing to 4G. Basically the protocol stack remains the same. Further to the protocol stack considerations in the SBA (chapter 2.4) the standardization of the infrastructure is moving away from niche signaling protocols that are specific to the telecom domain and focuses to a protocol that is widely supported in the IT industry today and has a significant development base. The evolution of the traditional mobile networks is reflected by the newly introduced mechanisms that the Network Repository Function provides (chapter 2.5). Another interesting concept is related with the API Exposure in 5G Networks (chapter 2.6) which potentially is used for troubleshooting purposes particularly in roaming scenarios. Then a comparison is performed between the standalone and non-standalone architecture (chapter 2.7) is presented in the terms of master and secondary carrier, core network choice, operator and vendor perspective and end user experience. The intelligence considerations chapter is ending with a network architecture comparison (chapter 2.8) of the 6 main pillars of the two mobile network generations, the RAN interface, the authentication mechanisms, the network slicing concept, the QOS model, the connection states and finally the cloud naiveness. Furthermore, by focusing to network slicing, a research regarding the diverse requirements of the different standardization organizations is presented (chapter 3). Continuing with a High-Level network slicing principles and mechanisms overview. 3GPP (chapter 3.1) is focusing on standardizing the mechanisms for the Network Slice Instance Selection and Association procedure which is thoroughly tackled in the terms of signaling flow. The Slice Association mechanism is binded with the security requirements defined and thus with the whole security infrastructure which is significantly adopted to that network slicing concept. The IETF (chapter 3.2) wants to describe a more comprehensive range of implemented end-to-end network slicing for both mobile networks, but also other networks modes which do not include both RAN and CN. GSMA (chapter 3.3) describes network slicing in a more generalized and business manner. ETSI (chapter 3.4) is focusing on standardizing the mechanisms for the virtualization of the physical network infrastructure resources aiming to define the network slicing architecture and thus the Network Slicing Lifecycle Management and Monitoring procedures. 3GPP is focusing on standardizing the mechanisms (chapter 4) among others and for the Network Slice Instance Selection and Association procedure which will be tackled in the terms of signaling flow (chapter 4.1). The Slice Association mechanism is binded with the security requirements defined and thus with the whole security infrastructure which is significantly adopted to that network slicing concept (chapter 4.3). Further to the approach followed by ETSI (chapter 5.1) 3 Core Layers of the Network Slice Architecture have been defined. The service instance layer, the network slice instance layer and the resource layer (chapter 5.2). Also 4 main phases of the network slicing lifecycle have been defined (chapter 5.3). The Preparation Phase, the Commissioning Phase, the Operation Phase, and the Decommissioning Phase. Further to this rational 3 architectural models have been intuitively defined reflecting the different use cases of network slicing (chapter 5.4). These are the Single Owner Single Controller Model, the Single Owner and Multiple Tenants Model, and the Multiple Owners or Multiple Tenants Model (chapter 5.5). The GSMA is focusing on standardizing network slicing and how it serves and promotes 5G roadmap but from a business perspective. GSMA focuses on the fact that different services should be provided concurrently via an evolved mobile network operating beyond the best effort model (chapter 6). Having addressed a thorough overview of the 3GPP, ETSI and GSMA perspective of network slicing, an architecture in the 5G communication system is proposed which basically bridges the mentioned mechanisms described. From a generic slice template defined by GSMA, is feeded into the OSS/BSS prototype called OpenSlice for a network slice to be instantiated monitored and configured on runtime with zero-day configurations under a shared infrastructure which relies in one or more administrative domains (chapter 7). Concluding with the following assumptions. Based on the above diverse definitions of network slicing, it is easy to find that the different organizations have a different interpretation of network slicing principle and have their focus on each one aspect (chapter 8.1). Referring to the IP perspective of network slicing, intuitively the concept of slicing a network into virtual containers or virtual elements is not new for IP networks. There are mechanisms already defined tackling the emerging requirements for the adoption of network slicing as a traffic engineering problem like VLANs, segment routing or MPLS TE (chapter 8.2). Also, the orthogonality of the functions applied in RAN leads to the introduction of the ORAN by disaggregating the RAN protocol stuck and introducing the SDN principles (chapter 8.3). Also, in order to achieve the requested quality of experience parameters defined, a disaggregation of the over-the-top applications into microservices is needed, in order to be able to be hosted in the mobile network cloud and thus closer to the end users. Slice security and slice privacy are two important aspects need to be considered by all Customer Serving Providers for the successful slice isolation (chapter 8.5). Summarizing, several aspects need to be considered from the Operation and Service Assurance Perspective for the Success of Network Slicing (chapter 8.6). Demanding needs like IOT and services requiring real-time interactions between components and additional emerging services overstretch the networks. Several considerations should be taking into account referring to management and operation procedures concerning also the network slicing architecture. Finally, infrastructure testing and scaling in the network maintenance should be considered from the Operators for the success of network slicing.

2. INTELLIGENCE CONSIDERATIONS IN 5G ARCHITECTURE

2.1 5G Evolution

Right now, 5G networks are being deployed and entering commercial service around the world. They will provide unparalleled connectivity to everyone and everything, everywhere. These first commercial deployment to 5G primarily focus on delivering enhanced mobile broadband using existing mobile frequencies and new millimeter wave frequencies which will open large amounts of new bandwidth. But these networks are only the beginning of the 5G Revolution. 3GPP which is the standardization partnership that produces the specifications for 5G is already working on the next phase of 5G known as release 16 [1] for which the specifications are scheduled to be completed in 2021. This would dramatically expand the ecosystem and the range of use cases that can benefit from 5G. It adds a set of features enabling ultra-Reliable, low latency and time sensitive Communications. These features are the key to supporting industrial IOT. The industrial internet-of-things which in turn is the foundation of the fourth Industrial revolution where all processes and machines and production environments are controlled wirelessly. Among other features Release 16 will also focus on introducing support for private network deployments including corresponding security features as well as enabling 5G to be deployed in unlicensed spectrum. This opens new opportunities for enterprises and campuses to unveil themselves of the benefits of 5G. The foundation provided from the first version of 5G is in fact so flexible that the sweeter features it contains will be able to be expanded in a backward compatible manner for many years to come. The vision for this continuing evolution of 5G is coming to fruition in early planning for Release 17 [2] targeted at the end of 2021. It will include enhanced support of industrial IOT with widest support for highly synchronous communications between devices. It is also expected a light version of 5G which would enable 5G devices is to be built at lower cost, with longer battery life while simultaneously providing higher data rates and reliability and lower latency than existing IOT technologies. As the 5G ecosystem continues to expand non-terrestrial connectivity will be integrated into the Networks providing coverage even in the remotest of areas by means of high-altitude platforms and lower orbit satellites. Release 17 will provide enhancements to network Edge Computing and Network Automation as well as further evolution of the all-important security features. Looking even further ahead the range of radio frequencies that can be exploited by 5G will be expanded with new technological developments opening up tens of gigahertz of new bandwidth in the range 52 to 115 gigahertz. Release 16 leads to the transformation of industries through digital automation with industrial IOT, unlicensed spectrum and private networks. The first 5G networks are already starting to change people's lives for the better. Over the coming years the range of new features that will be embraced by the continuing evolution of 5G will take human communication and industrial automation into a completely new realm bridging the physical and digital worlds to deliver the extraordinary.

2.2 State of the Art Concepts in 5G Architecture

2.2.1 Service Based Architecture

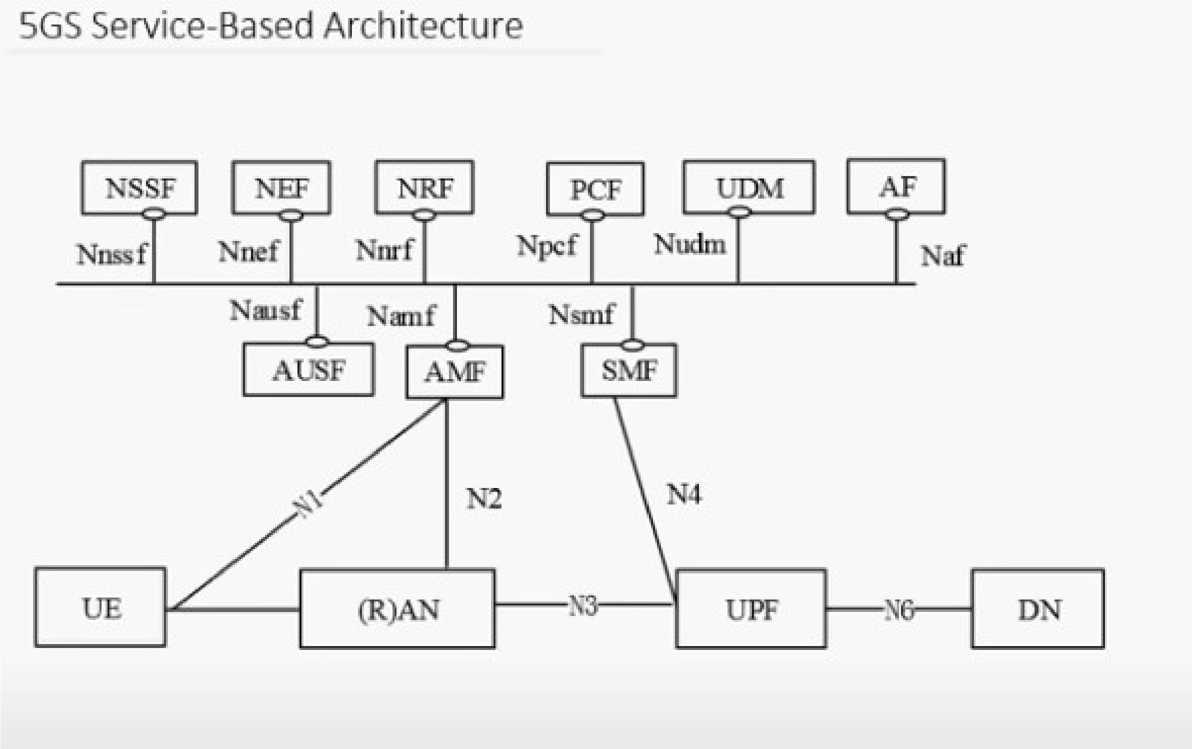


Figure 1: TS 23.501 SBA Architecture

One major Improvement in the infrastructure of the Core network Architecture introduced with 5G is the so-called Service Based Architecture (SBA) [3]. The SBA is based on the abstract concept of an application function. In contrast with 4G these peer-to-peer interfaces between any two elements has not been adopted in the Packet Core. Service based architecture is based on an infrastructure in a publisher and consumer model where any of the functions can publish their own APIs and any other function can go ahead and consume those APIs, by facilitating reusability off all these functions and making the standardization much easier. Mechanisms have been defined to accommodate a communication procedure through a common message bus, within the functions.

2.2.2 Orthogonality of the Functions

Another interesting concept in the design of 5G Core is the orthogonality of functions. Every function or every node defined in the 5G core performs only one type of function or only one type of management inside the core. That concept introduces two separated functions for the handling of mobility management and session management in 5G Core where the Access and Mobility Management Function (AMF) performs only mobility management. The handling of session management is the responsibility of the Session Management Function (SMF). So, in 5G Core there are two main functions responsible for the handling of the control plane signaling. Further to the orthogonality of the functions where every function serves for a specific purpose, in 5G Core is introduced a clear separation between control and user plane signaling. There is a separation in the interfaces that the gNB communicates with the Packet Core. Control Plane Signaling is forwarded through N2 Interface to the AMF and User Plane Signaling is forwarded through N3 interface to the User Plane function (UPF). In 5G Core there is only one user plane element. This user plane element can either be distributed in a way where the internet offload point will be closer to the user (possibly as content cache), or it can be a centralized user plane function. That centralized user plane serves for more like centralized peering Points where the UEs can have multiple IP anchor points across the network. The IP Pool that the UE holds and announces, gives the ability in the network to install some content caches and do local breakout for specific applications closer to the user.

2.2.3 SDN Principles Adoption

Another worth mentioning characteristic is that since the user plane functions and the session management functions have been completely separated, an SDN type of environment has been introduced. The Session Management Function (SMF) has a role of an SDN Controller for the user pane functions. Specific Signaling is forwarded through the N4 interface once the SMF needs to set up some flows or program some existing flows when the UE wants to access the internet or requesting an additional service.

2.3 Protocol Stack Considerations in Backhaul

From a back-hole perspective there is no significant difference comparing to 4G. Basically the protocol stack remains the same. The user plane uses GTP_U and the control plane uses SCTP. Differences can be found in the introduction of a new control protocol which is called NGAP, and in the X interface where X2 has not been adopted and has been replaced with Xn between the different radios for better coordination.

2.3.1 RAN Architectural Comparison with 4G

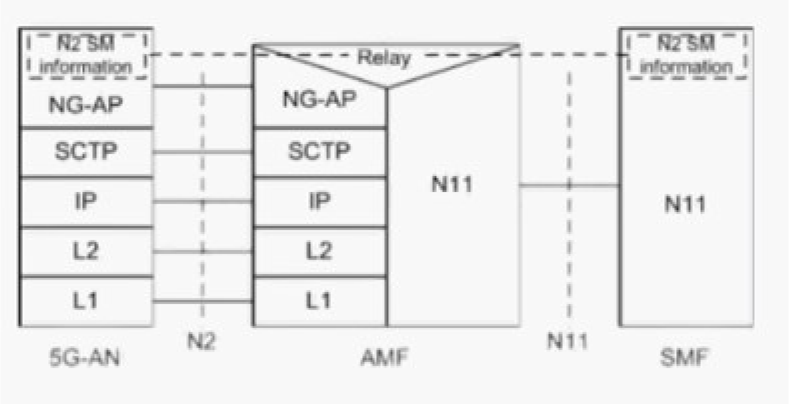


Figure 2: RAN Control Plane Signaling Stack

The new radio (NR) [4] is not part of the service-based architecture, but it belongs to the 5G Core. There is no HTTP 2 interface in the RAN. The RAN is the only part of the whole system that preserves the SCTP transfer protocol for OSI layer 4. This is the one and only exception to the whole architecture that uses SCTP. In Comparison with the Evolved Packet Core the EPS radio access bearer is extending as equivalent with the User Plane Function (UPF). From the radio perspective what used to be the SGW has been replaced with the plane function UPF. SGW used to talking to an MME whereas the MME is being replaced with an AMF. The S1 interface is basically replaced with N3 in the case of user plane Signaling and with N2 interface which was towards the MME is towards the AMF for the control plane Signaling. Referring to the protocol stack relative to the access network interfaces to the core for user plane does not change. Now for control plane it does change as the AMF replaces the MMEs role towards the radio network thus s1AP interface is replaced with the Next Generation application protocol (NGAP). From an architecture perspective this is not a major change in the protocol stack as S1AP and NGAP appear to be almost identical as they are both asn1 encoded binary sibling Protocols. Differences can be determined in Mobility management whereas the NAS and Mobility management Procedures have been implemented differently but there is a common terminology where ECGI becomes NRCGI for the Cell IDs and tracking area Identity remain identical.

2.4 Protocol Stack Considerations in SBA

3GPP decided that for the underlying signaling protocol in N4 interface GTPC will be reused by creating the Packet Forwarding and Control Protocol (PFCP). The PFCP main functionality is to send flow descriptors to the UPF and with those flow descriptors can define several actions depending on the attributes of every flow. The communication between the functions in the service-based packet core is based on restful interfaces and the underlying communication protocol is the HTTP2.

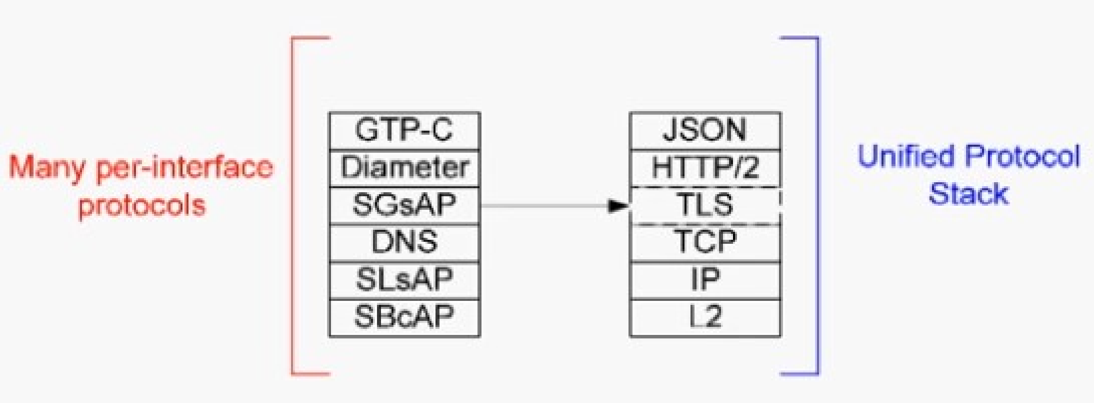


Figure 3: Protocol Stack Comparison between 4G and 5G

So, in favor of one HTTP2 signaling stack, all the other signaling protocols SBcAP, SLsAP, DNS, SCsAP, Diameter, GTP-C that existed in the evolved packet core and they are all being replaced with this single stack IP, TCP, TLS, HTTP2 and JSON. The standardization of the infrastructure is moving away from niche signaling protocols that are specific to the telecom domain and focuses to a protocol that is widely supported in the IT industry today and has a significant development base. One common interface for a network function will support any type of request for that Network function. Following this rational the induce of signaling protocols is performed right once rather than having five or six different protocols with each of their own advantages.

2.4.1 TLS Concept

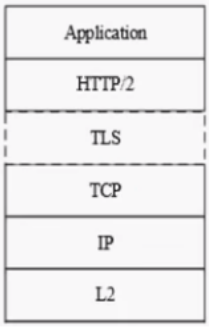


Figure 4: 5G Core Network Protocol Stack

The protocol stack defined by 3GPP introduces TLS encryption optionally and as the dotted line around TLS. In terms of practical vendors deployment could end up with unencrypted traffic within the network and encrypted between networks or TLS would be adopted for every procedure.

2.4.2 HTTP 2 Concept

HTTP2 Supports multi-stream applications. In the contrast with HTTP, supports multi-streaming over TCP so it is not affected from common HTTP vulnerabilities like header blocking and line header blocking. HTTP had a limit on how many requests could be in wait before getting responses for new requests can be made. With multi-streaming HTTP 2 basically takes a lot of SCTP Concepts and applies them on HTTP. Supports Pre-emptive data pushes whereas a response to a request will include the answer to what is being asked and the answer to what might be asked in the next request. That is something HTTP 2 supports which was previously not supported in HTTP. Moving forward with HTTP 2 all that is remaining is to define the application of intelligence in order to get the actual protocol stacks working with little effort, and in such a way that the effort will be shared with huge potential for synergy with the rest of the IT industry.

2.5 A State-of-the-Art Concept behind Network Repository Function (NRF)

The network repository function [5] concept is in essence replacing DNS with a much more scalable way for gateway selection essential for the 5G core. Network functions register themselves with their services and IP information in an entity called NRF. So, the idea would be rather than going to a computer and configuring DNS records for the Node as exactly happened in 4G, a 5G Core entity when it connects to the network is going to register itself with an edit in the network repository. This registration procedure includes information about the supported services, about the enabled interfaces (layer 2 information), IP related information bind to the supported services (layer 3 information). The registration procedure is taking place for all the functions in 5G Core in such a way that every function in the network could actually get and pull data from the network repository by gathering all the required information in order to initiate a 3GPP Procedure with another function rather than doing something specific to Gateway selection in DNS. This much broader network discovery concept that is possible between systems including their own data into the NRF and be retrieved by other systems that need to talk to them and to reduce the number of places where data needs to be defined. Also reduces the potential for human error and increases in the scope of what can be learned and can be done between systems in the core by replacing DNS with this repository concept.

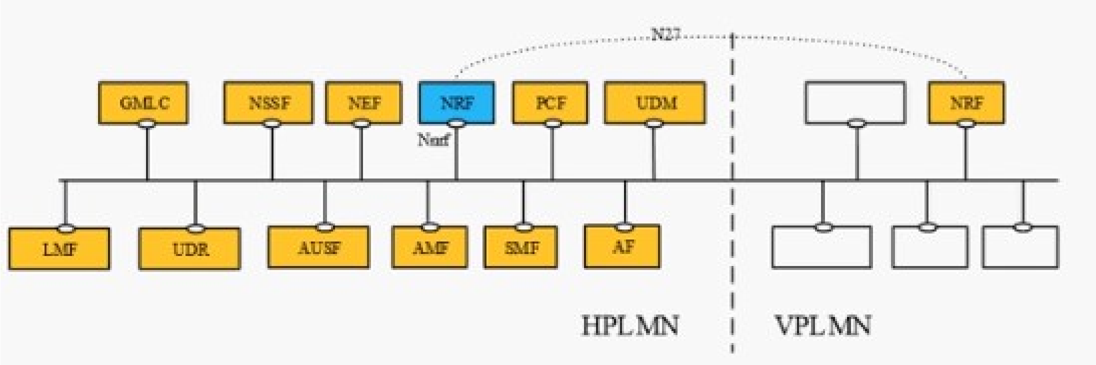


Figure 5: TS 29.510 NRF Functionality in Roaming Scenarios

An interesting similarity to the DNS in the evolved packet core is that the MCC and MNC that is a PLMN identity of an operator is used as the top level of its datastore hierarchy. That lends to roaming scenarios where an operators PLMN identity is uniquely defined and if an operator will be asked to be uniquely discovered then other operators have a unique point of reference to get to all of the operator's data. The NRF includes network function and slice related information required to determine the appropriate AMF and SMFs. The NRF is used to determine which slices are instantiated and available to the AMF. The NRF is configured and updated by the management plane whenever a new network slice is created or modified. The selected SMFs may also establish a PDU session if requested in the NAS message. The PDU session request may include the DNN for the corresponding SST/SD and the Session and Service Continuity (SSC) mode.

2.6 Exposure Capabilities of 5G Architecture

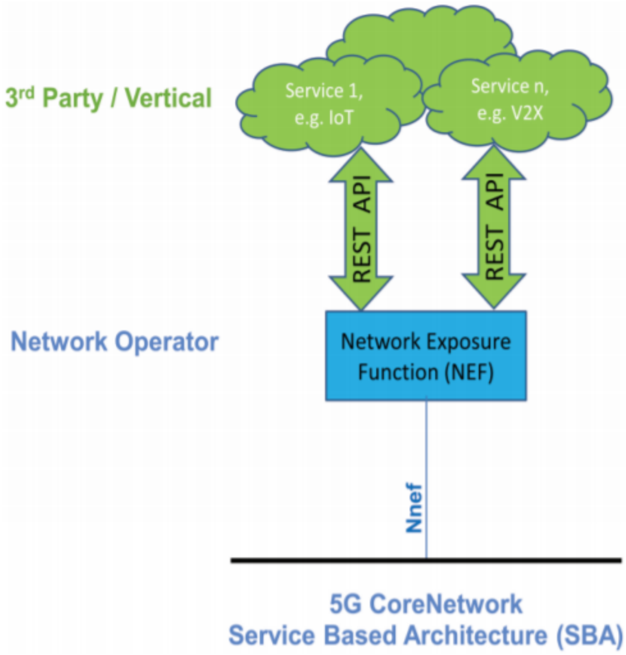
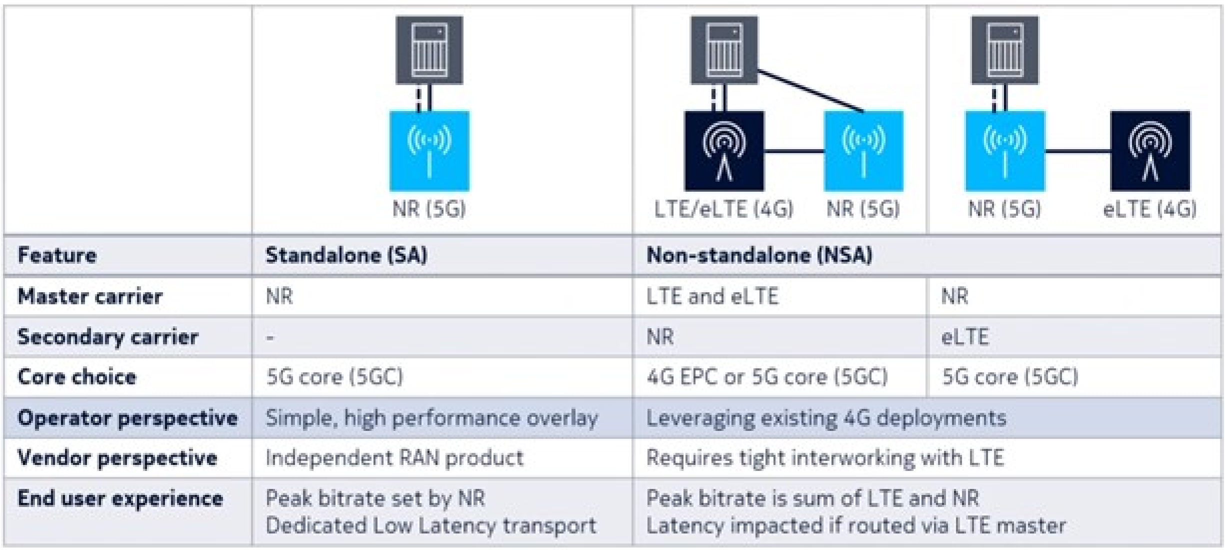


Figure 6: NEF (as part of the 5G SBA) providing services to 3rd parties via RESTful APIs.

For the API exposure concept [6], a new function has been introduced called as Network Exposure Function (NEF). NEF provides the capability of the network to make available the whole 5G system and its network functions within defined security levels and security hierarchies. It supports three types of operations. Handling monitor requests, subscribing and notified of state changes for sessions that exist in another network functions. It can provide provisioning requests to change data in network functions. Also, when it will be asked as a request type to invoke for a policy or charging change from externally, then the NEF requests policy and charging changes to the related function inside the Packet Core. It is potentially useful for analytic tools to subscribe for data or to pull data by giving access to more data from more systems and in a more generic way. That lends to Big Data and analytic tools and give more potential applications for them. Potentially is used for troubleshooting purposes particularly in roaming scenarios. It gives to an operator the functionality to get some information that cannot get manually in roaming scenarios. If an operator is trying to support a complaint of poor service for a roaming scenario there are advantages if the operator has an interface to request information from the serving operator about the subscriber status. Without the NEF such a scenario triggers a manual processing between the operators. NEF gives the ability to query and subscribe for data so within a given operators PLMN and based on the SLA agreements to grant permissions between operators to query for the session State and Mobility state of a subscriber and the session management state of a subscriber and thus can learn as the home operator about its own subscribers even as they roaming to other operators. Also, and through the subscription concept to the visitor operator, which was not something possible in the evolved packet core architecture, but more in generically the exposure concept allows for non 5G core systems to get information from the 5G core in a way that's not standardized today for anything really equivalent.

2.7 Stand-Alone (SA) and Non-Standalone (NSA) Architecture Comparison

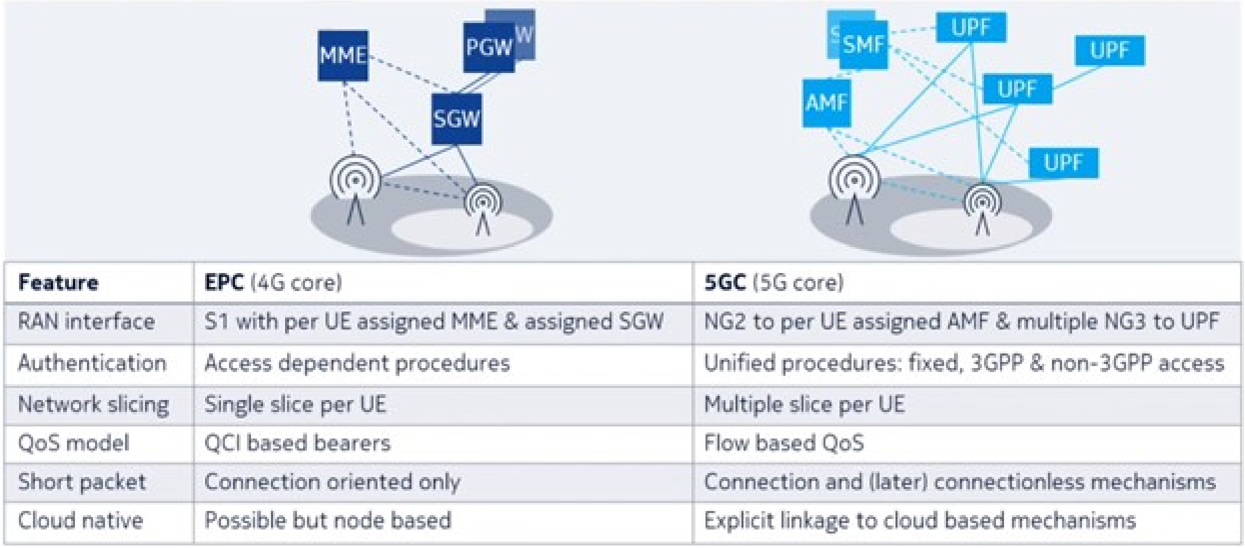

NR (5G)
LTE/eLTE (4G)
NR (5G)
NR (5G)
eLTE (4G)

Feature	Standalone (SA)	Non-standalone (NSA)	
Master carrier	NR	LTE and eLTE	NR
Secondary carrier	-	NR	eLTE
Core choice	5G core (5GC)	4G EPC or 5G core (5GC)	5G core (5GC)
Operator perspective	Simple, high performance overlay	Leveraging existing 4G deployments	
Vendor perspective	Independent RAN product	Requires tight interworking with LTE	
End user experience	Peak bitrate set by NR Dedicated Low Latency transport	Peak bitrate is sum of LTE and NR Latency impacted if routed via LTE master	

Figure 7: Standalone and Non-Standalone NR Functionality

The functionality in terms of the radio is divided in two architecture options [7][8] as there is a demanding need of the Operators for a smooth transition to 5G. In the first option called as the standalone option the New Radio (NR) is directly connected to the 5G core. There are two more options called the non-standalone options whereas the LTE is connected to the Core network and the NR connected to the LTE and vice versa. More Precisely, 5G NR is connected to the Core Network and the eNodeB is connected with the gNB and then connected to the core 5G Core. In the standalone option the master carrier is the new radio and in the non-standalone the master carrier is LTE or evolved LTE in the first option, and in the second one the master carrier is the NR. The meaning of master carrier is instantiated as the new radio cannot connect to the core directly for the signaling messages. It should be able to connect using the existing infrastructure of another radio or another RAT technology in order to connect with the core network. In the standalone option the NR does not need a secondary carrier. In the non-standalone options depending on the master carrier, the secondary carrier in the one option is the NR and in the other option is the evolved LTE. In terms of the core network in standalone option can be only 5G Core but in the non-standalone it might be the 4G EPC or the 5G Core by giving the capability for the operator to make a choice. The benefits for the operator in the standalone option is a simple and high-performance overlay by deploying directly the 5G Core without the need of an interworking mechanism for supporting the older generations. In contrast with the first non-standalone option where the existing 4G deployment must be leveraged in a manner where its software be able to support the secondary NR carrier and by introducing a tight interworking mechanism between the LTE and NR. From the engineering perspective there are some advancements in the handover procedures. The software capabilities need to be tightly coupled in order to support efficiently this interworking in terms of performance of latency and generally in terms of quality of service. Referring to the end user experience in the standalone option, the peak bit rate is set by the new radio from a dedicated low latency transport mechanism without any hops to LTE. In the non-standalone option, the peak bit rate depends only on the minimum between LTE bit rate and the 5G bit rate. Consequently, this combination of both has a direct impact to the latency. By having the LTE as the master carrier, a 5G UE is accessing the network and its applications are requesting for the 5G capable latency and QOE which is bound to what LTE is offering depending on its limitations. This is an important factor that needs to be considered when this non-standalone option needs to be deployed. In the standalone option whereas there is a direct connection to the 5G Core, and the latency is provided within the permissible required levels.

2.8 Network Architecture Comparison



Feature	EPC (4G core)	5GC (5G core)
RAN interface	S1 with per UE assigned MME & assigned SGW	NG2 to per UE assigned AMF & multiple NG3 to UPF
Authentication	Access dependent procedures	Unified procedures: fixed, 3GPP & non-3GPP access
Network slicing	Single slice per UE	Multiple slice per UE
QoS model	QCI based bearers	Flow based QoS
Short packet	Connection oriented only	Connection and (later) connectionless mechanisms
Cloud native	Possible but node based	Explicit linkage to cloud based mechanisms

Figure 8: Architecture Key Features Comparison 4G and 5G

The network architecture comparison between 4G and 5G networks can be summarized to the below interesting points [9][10]. The main elements in the LTE Core network are the MME for Mobility Management Entity, the SGW for Serving Gateway and the PGW as PDN Gateway or Packet Gateway. In the 5G Core network there is the AMF for Access and Mobility Management Function, the SMF for session management function and the UPF for User Plane Function. The data throughput, and the packet flows per slices are managed by the user plane function. In the 5G Core can exist multiple instances of UPFs or SMFs or AMFs by giving the capability to the network and depending on the slice instances to use different virtual function instances. Further to the differentiation between the 4G and the 5G, the 4G RAN interface instantiates a static end to end connection per UE, assigned to a specific MME and assigned to a specific gateway, the one that is closely coupling. In the 5G Core there is an AMF assigned per UE and multiple UPFs. The access network interface is the N2 interface transferring the control plane traffic between the new radio and the AMF. The N3 interface is used for transferring the user plane traffic between the new radio and the UPF. Multiple UPFs may serve on UE depending on the allowed slices. One UPF for the mobile broadband, another one for the IOT slices, another for the ultra-low latency communications and another for enterprise solutions. A UE with specific capabilities located under specific cells can select any of the UPFs and more than one. This architecture is not supported in LTE as the MME always selects a specific serving gateway and that is considered as a restriction. Referring to the authentication procedures in 4G the procedures are access dependent based to the UE type whereas in 5G Core unified access like fixed, 3GPP or non-3GPP (Wi-Fi Connections) is supported using the same authentication mechanisms. The 5G procedures are unified and the access and mobility function is responsible for the alignment and the coordination of the RAN part. Quality of Service is another major difference between the 2 architectures where in 4G is based on the QOS Class Identified QCI based bearer. Depending on the dedicated bearer, a specific QOS is assigned in IP Flows and thus a specific QCI value. In the 5G Core the QOS is assigned depending on the resources required for the flow (flow-based QOS). The resources needed for every flow like bandwidth is assigned based on the type of service. This kind of dynamic connectivity is not supported in 4G where irrespectively of the type of service and the required resources, the same QCI value will be assigned. The architecture of 4G defines that the UE should be always connected. Therefore, the underlying signaling consists of short packet lengths but in a not at all battery efficient way. The underlying mechanisms offered by 5G support different UE states where the UE can be registered and in an idle state. Referring to the cloud naiveness the static 4G architecture can be instantiated either from physical nodes or from virtual Nodes. Inherently to 5G is introduced a dynamic network architecture with a non-constant number of elements. A demanding need is emerging along this rational for cloud native functions in order for the 5G Architecture to be supported as it is explicitly linked to cloud based mechanisms. All the 5G network elements are only based on the cloud comparing with the 4G architecture where they usually instantiated as physical nodes.

3. NETWORK SLICING RESEARCH BASED ON DIFFERENT REQUIREMENTS

As one of the important innovations in 5G, there are diverse business communities and standard organizations dedicated to research Network Slicing, such as the IETF, 3GPP, GSMA, and ETSI. The organizations and communities not only standardize network slicing specifications, but they also independently concentrate on defining network slicing in their own way. Aiming at the complete standardization of network slicing, there are variations of specifications between different organizations and communities. The following chapters illustrate that network slicing has been established by several Standards Definition Organizations (SDOs) and communities. However, the meaning and definition of each regarding the network slicing operation is different. There is no common definition of network slice and it’s provided services to the vertical markets. There are variations between each organization based on opposing views on network slicing mainly because every company is focusing to a specific aspect of network slicing mechanisms e.g., marketing aspect, architecture aspects. The 3GPP has released network slicing standard concerns how architecture is defined [3] and in [11] concerns management and orchestration. The 3GPP specifications provide the criteria, use cases and other related functionality of network slicing. Due to its significance among recent network technologies, the topic of network slicing has been widely discussed in the IETF. Besides, there are other research organizations doing the same form of work in a related area, such as the European Telecommunications Standards Institute (ETSI), the Broadband Forum (BBF), and the GSM Association (GSMA). Each describes the various field of considerations for slicing networks differently.

3.1 3GPP

3.1.1 Definition of Network Slicing

Recognizing network slicing as an essential aspect of 5G, 3GPP committed in creating a comprehensive specification of network slicing principles from the very early period since the beginning of the discussion on 5G technologies. 3GPP only discuss the architecture and management of Network slicing in 3GPP, which are mostly related to [3] and [11]. 3GPP describes the network slice as a logical communication network consisting of the core network (CN), the user plane network functions (UPFs) and the Access Network (AN). One User Equipment can be served by several slices concurrently via a single RAN. Each network slice within the PLMN can differ based on the S-NSSAIs with different Slice/Service Types. Therefore, consumers may have a preference of various internet service providers to have the same service quality. The S-NSSAI, provides identification and selection information of a network slice, comprised of a Slice/Service Type (SST) which refers to the desired Network Slice behavior in terms of features and services and a Slice Differentiator (SD) which is optional information that complements the Slice/Service Type(s) to distinguish between multiple Network Slices of the same Slice/Service Type. A S-NSSAI may have one or more Network Slice instances connected with it. The SST value shows that the 3 service categories include Enhanced Mobile Broadband (eMBB), Ultra Low Latency Communications (URLLC) and Massive IoT.

3.1.2 Network Slice Operation Procedure

Network slicing association is performed in UE registration, in which the UE is requesting for a network slice based on the first contacted access and mobility management function (AMF). After AMF is chosen by the RAN, receives the UE Registration request, the UE sends a requested NSSAI to AMF, and AMF is required to determine if it can serve the UE by checking whether the S-NSSAI(s) in the Requested NSSAI are permitted based on the Subscribed S-NSSAIs. In the case is permitted, AMF query the Network Slice Selection Function (NSSF), with requested NSSAI, mapping of requested NSSAI to configured NSSAI. The NSSF returns to the current AMF the Allowed NSSAI for the appropriate access types and the UE, after receiving an allowed NSSAI from the serving AMF, will store it. As one of the most essential goals of network slicing architecture, is to create and maintain a PDU session in a Network Slice. A PDU Session enables data transfer in a Network Slice. Data transfer begins after a PDU session to a data network is created. Each PDU Session refers to one core network slice and one radio access network slice explicitly.

3.2 IETF

3.2.1 Definition of Network Slicing

The architecture of network slicing is addressed in terms of its terminology, architecture [12], use cases [13], problem-solving, and other related aspects. In order to determine the concrete technical specifications of network slicing offerings, a technology-independent information model is created. In order to achieve the objective of bridging top-down and bottom-up approaches to a technology-independent management plane, COMS (Common Operation and Management on Network Slices) is defined and explained here [12]. The word "end to end network slice" in [14] means the cross-domain network slice that includes network elements of the access network, the transport network, and the core network. The definition of network slice is extended to include many functional components which shape numerous networks. To accommodate the multiple requirements of users, network slice instances can include many components which each require a particular amount of network resources and attributes.

3.2.2 Definition of Network Slicing Management and Orchestration

Network slicing Management and Orchestration is the process which produces the network slice instances, and maps resources to the requirements and capabilities of the users. The Characteristics of network slicing include anonymity, discretion, stability, versatility, sustainability and scalability. The Network Slicing Management and Orchestration is comprised of:

1. Template Management: A comprehensive overview of the specification and setup of the network slice instance and how the instance was instantiated, monitored and operated over the life cycle.
2. NS Repository: A set of policies defining how network slices are chosen and how instances are paired with network slices.
3. Life cycle Management and Monitoring: Network slicing and slice instance must be handled over the life cycle and monitored for problems.
4. E2E Orchestration: E2E Slices Orchestration and its features which can be autonomously organizing a range of interrelated resources, autonomously controlling slice life cycle management, autonomously coordinating and activating slice elasticity and logical resources positioning in slices, autonomously coordinating and re-configuring logical resources.
5. Domain Orchestration: Where several domain techniques and technologies are integrated, to orchestrate more efficiently, with less complexity and in an automated way the different domains.
6. NS Manager: The manager handles all access permissions and all interaction with each network slice and any external functions unique to that network slice instance.
7. Resource Registration: Responsible for controlling the exposure of network capabilities and states.

3.2.3 Information Model

However, in some situations, the verticals or the tenants of the offered networking capabilities and systems are interesting for a network-agnostic interface. A solution to this is an in-demand standardized information model that bridges the breach between technology-agnostic network slicing service requirements and actual implementations from the vendors. The information model for slicing the network includes the necessary capabilities to define the entities that constitute the slice and how they execute their delegated roles. The information model includes the connectivity of the out-of-date technologies, and a simple overview of the network slices entities and technologies across different domains.

3.3 GSMA

GSMA gives an overview on network slicing and how it serves and promotes 5G roadmap but from a business perspective in its document [16]. GSMA explain the idea of network slicing from the perspective of the vertical companies, which is the embodiment of the concept of running multiple logical networks as virtually autonomous business processes on a single physical infrastructure in an effective and economical manner. They specify the dynamic network capabilities include data speed, latency, efficiency, reliability and security. Also, in GSMA definition is mentioned that various operators may share the same network slice.

3.4 ETSI

ETSI advises how to map network slicing use cases identified in other SDOs to the ETSI NFV architecture framework [15] by specifying the underlying mechanisms when mapping the use cases to the NFV architecture. Instead of specifying the standards of network slicing, ETSI focuses on the mapping of the network slicing to NFV definition and explain how NFV architecture can enable it within the requirements of diverse organization such as 3GPP. Under the 3GPP concept of network slicing, ETSI defines NFV-NS (Virtualized Network Resources for the slice subnet and access to the physical resources) In addition, ETSI defines the Os-Ma-NFvo interface point that is involved in the communication between Slicing Management and NFV-MANO, as well as a bunch of management decisions from NSMF and NSSMF.

4. THE 3GPP PERSPECTIVE OF NETWORK SLICING

4.1 High-Level Network Slicing Principles and Mechanisms Overview

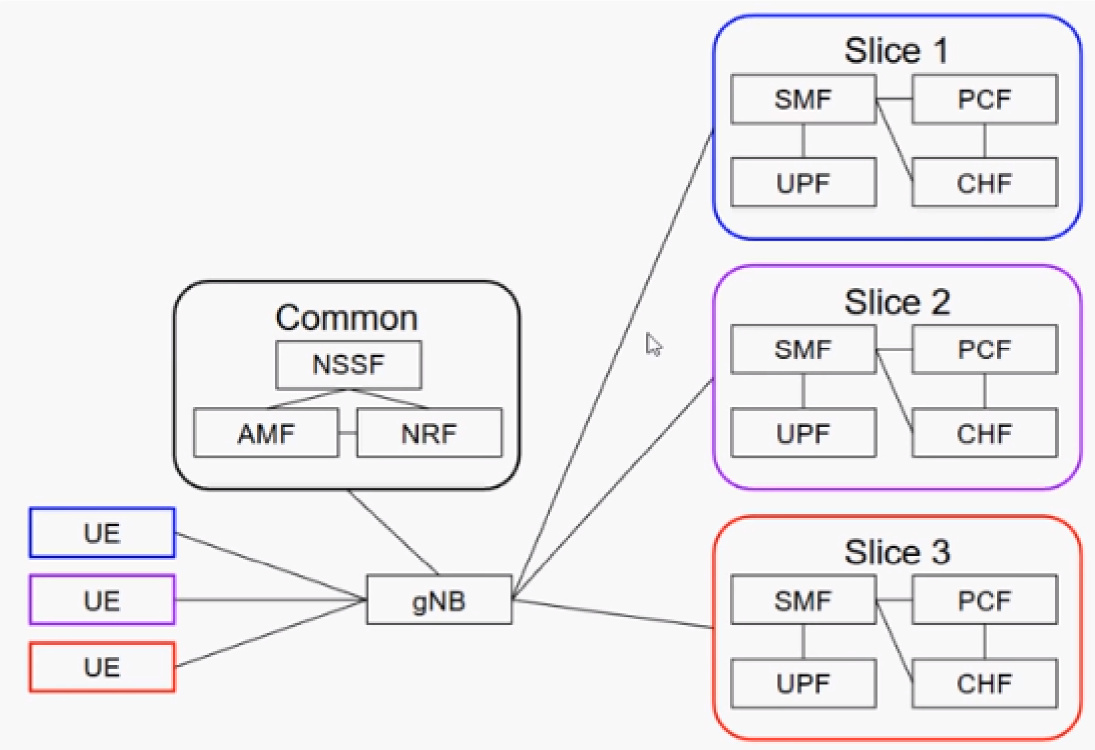


Figure 9: High Level Network Slice Overview

From a high-level architecture perspective network slicing can be considered as a session management profile which should invoke several mechanisms, some of them offered for policy control, some others for session management and some for charging. Generally, can be considered as a service profile where a single network is logically divided into slices of networks to provide a different type of service to specific subscribers. According to the 3GPP every network function which can be considered as 3GPP adopted or defined processing function in a network, is a Virtualized Network Function (VNF). A network slice is a logical network that provides specific network capabilities and network characteristics. A network slice instance is a set of network function instances and the required compute, storage, networking resources which form an End-to-End deployed network slice. According to 3GPP these are communication services, combined with several core networks and with several access networks. 3GPP defines that for a given service a configurable number of core network instances and a configurable number of access Networks. In service-based core network architecture some of the network functions are going to be device-specific. These would be the Access and Mobility Function (AMF), the Network Slice Selection Function

(NSSF) and the Network Repository Function (NRF). In contrast with Session Management Functions (SMF)s and User Plane Functions (UPF)s could be slice specific. Also, policy Control Function and the charging function and possibly different user plane functions can be used between the different network slices. Every slice can be illustrated virtually so virtual PCFs gets invoked in a network slice but not gets involved in another network slice. So physically in the 5G Packet Core might exist a physical PCF and multiple virtual PCFs which treat requests differently based on the network slice. Intuitively a slice can be considered as a profile ID that maps all the logic should be provided for a subscriber.

4.2 Network Slice Instance Selection and Association

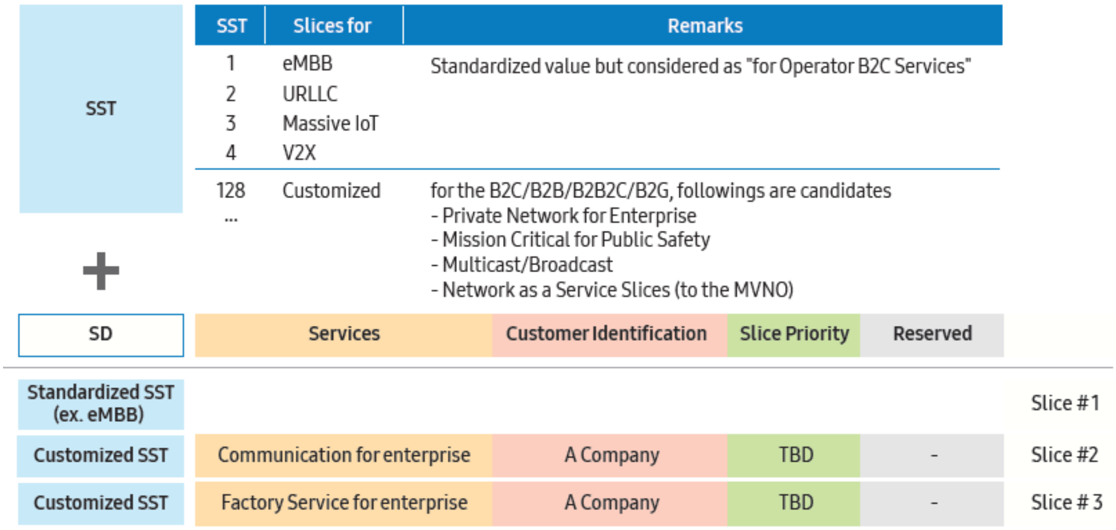
4.2.1 High Level Description

A UE can register for up to eight parallel slices[17]. So instantaneously up to eight UPFs might forwarding user plane traffic to UE. Each of these services are a-priori composed by the operator and each one has an actual numbering formatted and represented via Single Network Slice Selection Assistance Identity (S_NSSAI) and is the profile number for network slice. The formatting similar to the old QCI is based on specific slice terminologies. The S-NSSAI identifies each network slice service and provides information to properly assign network slice/functions. An S-NSSAI is comprised of:

•A Slice/Service type (SST), which refers to the expected network slice behavior in terms of features and services.

·A Slice Differentiator (SD), which is an optional information that complements the Slice/Service type(s) to differentiate amongst multiple network slices of the same Slice/Service type.

3GPP allows the use of the Slice Differentiator (SD) field that can build customized network slices. The SD field can be used to describe services, customer information and priority. Figure below illustrates the use of SST and SD.



SST

SST	Slices for	Remarks
1	eMBB	Standardized value but considered as "for Operator B2C Services"
2	URLLC	
3	Massive IoT	
4	V2X	
128 ...	Customized	for the B2C/B2B/B2B2C/B2G, followings are candidates - Private Network for Enterprise - Mission Critical for Public Safety - Multicast/Broadcast - Network as a Service Slices (to the MVNO)

+

SD	Services	Customer Identification	Slice Priority	Reserved	
Standardized SST (ex. eMBB)					Slice #1
Customized SST	Communication for enterprise	A Company	TBD	-	Slice #2
Customized SST	Factory Service for enterprise	A Company	TBD	-	Slice # 3

Figure 10: S-NSSAI Structure

The S_NSSAI SST of type 1 corresponds to Enhanced Mobile Broadband (eMBB), the SST of type 2 corresponds to (Ultra Reliable Low Latency Communications) URLLC and the SST of type 3 to (Massive Internet of Things) MIot. So, a UE during the registration procedure is sending a request to the AMF with up to eight slices in it. The AMF receives the request and validates that the slices requested by the UE are valid to be provisioned based on the subscription SLA. After the Successful validation, the request is propagated to the Network Slice Selection Function (NSSF) to inform AMF about which slice instance should access in order to connect to the PLMN as well. The NSSF responds with the slice instance that corresponds to UE. Then the AMF asks the network repository function about all the required information needed to derive which NFVs corresponds to the required network slice instance to compose the End-to-End connectivity.

4.2.2 Network Entities Involving

The Elements involved in the registration procedure are the user equipment, the Radio Access Network (RAN), the Access Mobility Function (AMF), the Session Management Function (SMF), the Policy Control Function (PCF), the Authorization User Subscription Function (AUSF) and the Unified Data Mediation Function (UDM). The registration procedure includes the Access and Mobility Function (AMF) selection procedure and the Session Management Function (SMF) selection procedure.

4.2.3 Types of Registration Procedure

The UE performs the registration procedure upon activated or as it is transitioned from 4G coverage to 5G coverage. There are four types, the registration procedure is divided into. The first type is referred to the registration performed when the UE is turned on, known as initial registration. The second is the periodic registration procedure similar to 4G tracking area update procedure. In this scenario after a specified amount of time determined by a timer assigned from the AMF to this UE, as part of the initial registration procedure, indicating when is necessary to regularly renew its registration with the AMF. The third one is the mobility registration procedure, which is similar to the periodic registration, but this would happen when a Subscriber moves from one tracking area to another other tracking area. The last case which could initiate the registration procedure would be in the event of emergency registration. This is often used by the UE when asking for emergency services in the NGRAN.

4.2.4 Network Registration and Slice Selection Procedure

The UE has to register with the network in order to be able to accept services, to facilitate mobility monitoring and to enable reachability [18]. The Registration procedure is used when whenever the UE has to initially connect to a 5G network or upon mobility procedure when the UE switches to a new Tracking Area (TA) in idle mode and when the UE performs a periodic update (due to a fixed timeframe of inactivity), etc. The 5G Core network may have already commissioned one or more slices in the network. The UE can be concurrently linked to several network slices. The management plane of an operator’s Slicing Architecture deployment shall specifically adhere to the following restrictions.

- The S-NSSAI (i.e., slice) must be supported in all cells in a UE’s registration area.
- Change in supported NSSAI for Tracking Areas (TAs) induces UE re-registration. A UE should be able to attach to multiple network slices with a single AMF.
- AS a UE is able to register to up to 8 network slices, AMFs may be common to a set of network slices.
- Further to the isolations between the slices the slice specific functions (e.g., SMFs and UPFs) should not be shared, and instead each one should be dedicated to a specific slice.

4.2.5 Registration Request

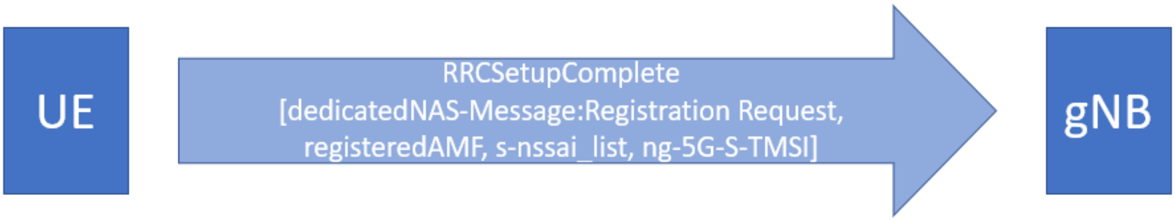

UE
RRCSetupComplete
[dedicatedNAS-Message:Registration Request,
registeredAMF, s-nssai_list, ng-5G-S-TMSI]
gNB

Figure 11: Registration Request

In the beginning, a 5G Device is enabled and an RRC signaling link between the UE and the RAN is created. The UE sends a NAS registration request via the RRC Setup Complete Message in the dedicated NAS-Message field. The registration request contains certain key attributes, such as mandatory information elements including security criteria or the registration type, with values based on either the initial registration or the periodic or the mobility registration update or the emergency registration. It would also have a mobile identity IE that holds temporary IDs such as TMSI or 5G-GUTI and thus a globally unique AMF ID or Subscriber SUCI. Mobile ID also carries permanent IDs such as SUPI in very unusual situations. Registration request can optionally include IEs like the requested Network Slice Selection Assistance Information (Requested NSSAI) or the PDU session status IE showing the available PDU sessions in the UE. The NSSAI defines the type of slice/service requested by the UE. Since the UE is not actually aware of the configuration of the network slice used by the operator when is connecting to the gNB. Thus, the UE may indicate the type of service/slice (SST) and the service differentiator (SD) is able to support in the registration request. The UE does not initially demand that all supported service categories be configured during the initial attachment procedure. In this case, the UE can determine which service types are initially needed by specifying the requested SST/SD in the registration request. The Radio Resource Control (RRC) message contains ng-5G-S-TMSI if allocated. The RRC message can also include additional details that may allow the RAN to choose an AMF for initial registration. The Additional details may include a pre-configured registered AMF ID or a pre-configured s-nssai-List. This information is used by the RAN to pick an AMF or AMF pool to serve the UE during the initial registration procedure. The NSSAI in RRC is intuitively a preconfigured entity, and the NSSAI in NAS may be referred to as the Configured or Accepted set of S-NSSAI. The RAN uses the information provided by the UE in the Radio Resource Control (RRC) message to decide which Access and Mobility Function (AMF) to choose for the UE.

4.2.6 AMF Selection and Relocation

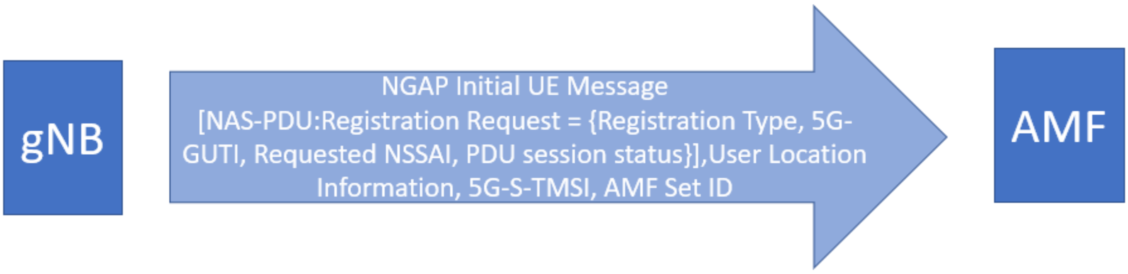

gNB
NGAP Initial UE Message
[NAS-PDU:Registration Request = {Registration Type, 5G-GUTI, Requested NSSAI, PDU session status}],User Location Information, 5G-S-TMSI, AMF Set ID
AMF

Figure 12: Initial UE Message / NAS Registration Request

4.2.6.1 AMF Relocation Optional Procedure

The AMF relocation procedure may initiate in parallel with the registration Procedure if the candidate AMF is not able to serve the UE. An AMF is getting a registration request and may need to reroute the registration request (due to slicing) to another AMF as the first AMF is not possibly the appropriate to serve the UE. The Registration with AMF relocation procedure re-routes the NAS message of the UE to a serving AMF during the registration procedure. As soon as the AMF is instantiated it registers its capabilities and some of the offered services at the NRF. Initially, the RAN sends a NGAP NAS message container message (the InitialUEMessage) to an AMF which carries the NAS registration request. In this case, if 5G GUTI s not valid and no S-NSSAI included in the NAS Message or the S-NSSAI(s) in the Requested NSSAI are not permitted according to the subscribed S-NSSAIs then the candidate AMF is the default AMF. Otherwise, the AMF is the appropriate serving AMF. In case of a default AMF, which is not the appropriate target AMF (as not capable of serve the UE) then the default AMF decides to reroute the NAS message to another, the serving AMF. For that purpose, the default AMF sends an NF discovery request, which includes the NSSAI, to the NRF to find a proper serving AMF that has the required capabilities to serve the UE. The NRF based on the information about the registered NFs and required capabilities in conjunction with the NSSF selects a serving AMF and transmits a NF discovery response identifying that AMF. The Reroute NAS message includes the information about N2 originating point for RAN the RAN UE NGAP ID, and the NAS message. Due to network slicing limitations and based on local policy and subscription information, the default AMF may determine to forward the NAS message to the target AMF via the RAN. In such a scenario the default AMF sends a Reroute NAS message to the RAN including the information about the serving AMF and the NAS Registration Request. If network slicing is used and the default AMF updates the NSSAI, the updated NSSAI is included in the redirection of NAS message. So based on the NSSAI in Reroute NAS message, the RAN sends the Initial UE message to the serving AMF. After a Successful Registration procedure, the NAS Registration Accept Message is carried by an NGAP Initial Context Setup Request message including the N2 terminating point for target AMF the AMF UE NGAP ID. The Procedure is completed with a successful outcome when the serving AMF receives Initial UE context setup response.

4.2.6.2 AMF Selection

The gNB receives a registration request from the UE and selects an AMF with the necessary capabilities. The gNB is connected to all AMFs within an AMF Region, via a N2-active connection. So, the gNB becomes aware of the capabilities of the AMFs during the N2 setup procedure. In a Scenario that the gNB is not able to select an AMF due to missing (optional) information from the RRC procedure, the registration request will be forwarded to a Default AMF for the AMF selection. So, the AMF Selection either is performed from the RAN based on the additional information provided in the RRC Procedure or the registration request will be forwarded to a default AMF for the AMF selection. There is only one default AMF for each gNB. There are two options for the default AMF: The standalone or default AMF is only responsible for picking a serving AMF and forwarding the request that he receives to the selected serving AMF. It is an isolated function that is not part of any slice (Anchor AMF). It is able to communicate with other AMFs. The purpose of isolating the slices is to prevent interfacing with the SMFs. Alternatively, often the AMF may be a default serving AMF that may also perform the AMF selection procedure. The Selection of the AMF in a case that the AMF relocation is required can be based on a couple of key points. The AMF verifies whether the S-NSSAI(s) in the Requested NSSAI are permitted based on the Subscribed S-NSSAIs and checks whether it can serve the UE. Then the default AMF performs a validation of a mapping between the requested network slices and the configured network slices in the current registration area. This default AMF possibly is not also a serving AMF. So, in cases the default AMF can’t serve the UE or if the UE context in the AMF does not yet include an Allowed NSSAI, the AMF queries the NSSF, with Requested NSSAI, the Subscribed S-NSSAIs, PLMN ID of the SUPI, location information, and possibly access technology being used by the UE. The NSSF selects the NSIs to serve the UE and determines the NRF(s) to be used to select NFs/services within the selected NSI(s). The NSSF thus providing one or more allowed network slices for the device and works with the NRF to determine the AMF Set for the requested Network Slice Instance. Under certain conditions where the AMF selection based on the NSSAI is not possible because either the AMF does not have the knowledge or R-NSSAI is not present, then the AMF selection is purely determined by the Subscribers 5G-GUTI. 5G-GUTI is composed of a set of AMF Identifiers or AMF ID which is composed of a Region ID, a Set ID and an AMF Pointer and 5G T-MSI. If the UE sends this GUTI as part of the registration request, the default AMF will be able to identify which was the last serving AMF by the use of AMF ID. Furthermore, can identify an AMF pool, where any AMF within the AMF pool can be used to serve the UE. If the RAN that the UE is accessing has an active connection with that AMF, that AMF will serve the UE otherwise the default AMF will be requesting from the old AMF the UE Context in order to serve the UE. There is also the initial attach scenario where 5G GUTI is not available and thus UE sends its SUCI without including the R-NSSAI. In Such case the default network slice for the UE is assigned, which is obtained from the subscription information. So, in scenarios where 5G-GUTI and R-NSSAI are not present the request is forwarded based on RAT to the default AMF which has been preconfigured on the new radio and then the default AMF can fetch the context from the old AMF. Once the serving AMF is selected for the UE, the serving AMF assigns a TMSI for the UE. The TMSI is used by the UE for subsequent Registration requests. The RAN transmits the Registration Request to the selected as new AMF via an N2 NAS Container message (depending on the registration type) which carries the NAS Registration Request and N2 parameters like the User location Information (containing the NR Cell ID and the tracking area identity) that the UE is located.

4.2.7 UE Context Transfer from the Old AMF

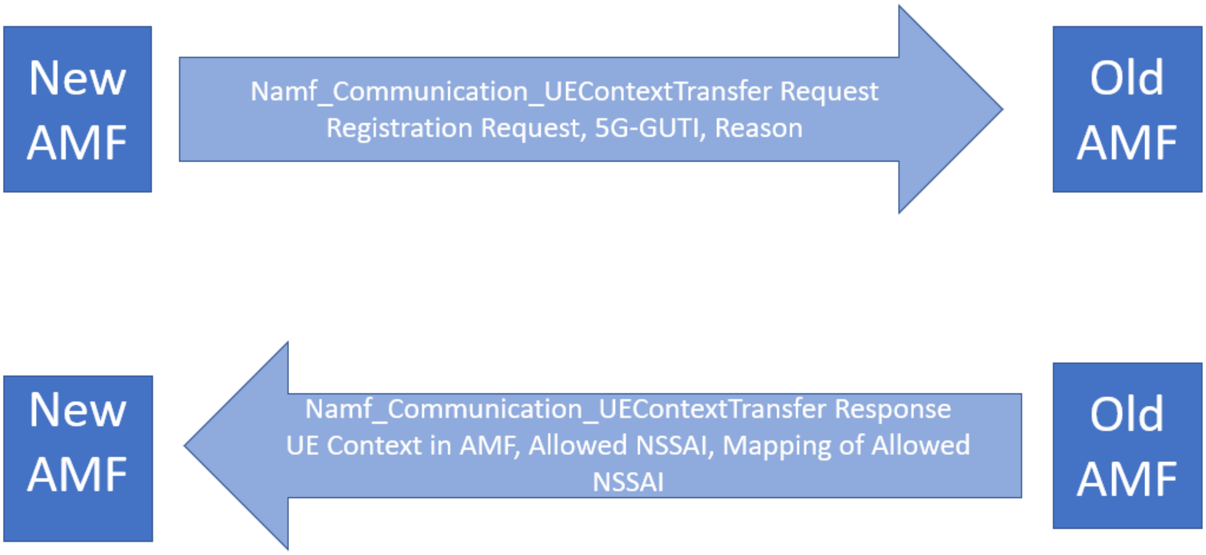

New AMF
Namf_Communication_UEContextTransfer Request
Registration Request, 5G-GUTI, Reason
Old AMF
New AMF
Namf_Communication_UEContextTransfer Response
UE Context in AMF, Allowed NSSAI, Mapping of Allowed NSSAI
Old AMF

Figure 13: UE Context Transfer from the Old AMF

Assuming that the new AMF is able to support the requested network slice assistance information, by examining the request decides if this UE was previously served by an old AMF. So, the selection of a new AMF is maybe initiated when the UE registers in an area where it is not served by the old AMF and the UE provides a globally unique AMF identifier containing the identity of the old AMF. The way that the new AMF retrieves the UE context from the old AMF is by using the 5G GUTI information element containing the AMF ID. A communication is being established by using the service Namf communications and then the UE context transfer service operation within this service. 5G-GUTI along with the reason to fetch the context and the registration request that UE sent to the network are composed to the **Namf_Communication_UEContextTransfer Request** message with recipient the old AMF. After an integrity check the old AMF is responding by sending an **Namf_Communication_UEContextTransfer Response** which carries the UE Context. The new AMF saves the UE Context obtained by the old AMF. The received UE Context includes the permanent identities as well as temporary Identities like the 5G-GUTI, the Subscriber Permanent Identifier SUPI, the Permanent Equipment Identifier (PEI), a bunch of DRX information and then all the subscription related information. So, a list of event subscriptions by other control plane network functions that had subscribed to AMF, indicating the events being subscribed as well as any information on how to send the corresponding notifications to any network functions that had subscribed for events such as IP address change or mobility for this given user. If the old AMF holds information about active PDU Sessions, the old AMF may include SMF information including SMF identities and PDU session identities.

4.2.8 Authentication and Security and Identification Procedures

The Authentication and the Security procedures are optional to the network operator as the UE Context is already created and is switching from one AMF which was the old one to the new one. The AMF transmits an Identity Request to the UE. If the SUPI is not provided by the UE nor retrieved from the old AMF, the Identity Request procedure is initiated by the AMF sending an Identity Request message to the UE including the requesting identity type. Usually, the requested identity type is the Subscriber Unique Concealed Identity which is the Concealed SUPI. The UE by using the PLMN public key is constructing it's SUCI and is responding to the AMF via an Identity response. Then the AMF performs a mapping between the SUCI and the SUPI and based on the resulting SUPI, selects an AUSF. The new AMF shall initiate authentication of the UE and NAS security procedures. The procedure of the AMF relocation may be initiated due to network slicing after the authentication and security procedures. If the authentication/security procedure fails, then the Registration shall be rejected, and the new AMF sends a reject indication to the old AMF. The old AMF continues as if the **Namf_Communication_UEContextTransfer** was never received. If the PEI was not provided by the UE nor retrieved from the old AMF, the Identity Request procedure is initiated by AMF sending an Identity Request message to the UE to retrieve the PEI. Optionally the AMF may initiate ME identity check with the EIR.

4.2.9 UE Context Transfer Acknowledgement from the Old AMF

If the AMF has changed, the new AMF transmits to the old AMF an Acknowledgement message **Namf_Communication_RegistrationComplete_Notify** that acknowledges the transfer of UE MM context.

4.2.10 User Connection Management Registration Procedure

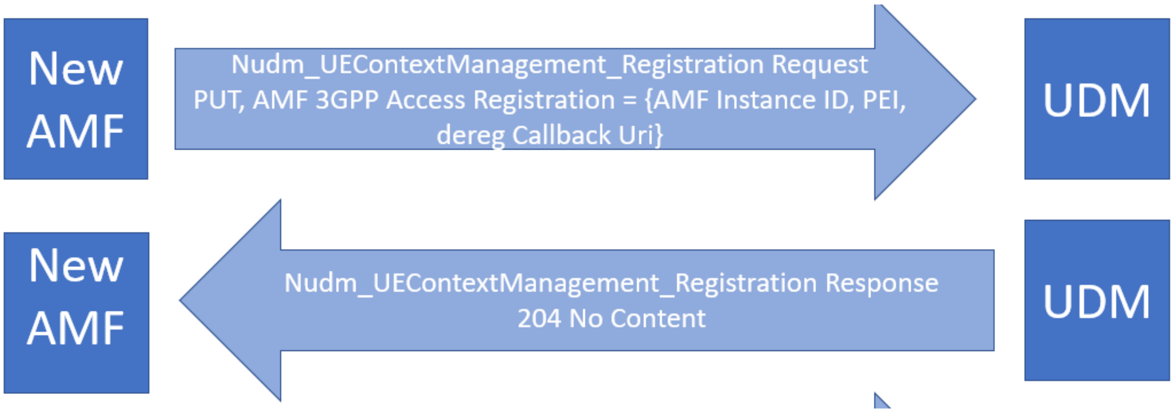

New AMF
Nudm_UEContextManagement_Registration Request
PUT, AMF 3GPP Access Registration = {AMF Instance ID, PEI, dereg Callback Uri}
UDM
New AMF
Nudm_UEContextManagement_Registration Response
204 No Content
UDM

Figure 14: UE Context Management Registration Service

At that point the user connection management registration process is initiated. Since the AMF has changed the new AMF shall register as the serving AMF for the UE in the specific access technology. The new AMF Selects a UDM based on the SUPI. The Procedure is initiated in cases the AMF has changed since the last registration, or if there is no valid subscription context for the UE in the AMF, or if the UE provides a SUPI which doesn't refer to a valid context in the AMF. The new AMF is sending to the selected UDM a Nudm**_UEContextManagement_Registration Request.** A response code "204 No Content" via the **Nudm_UEContextManagement_Registration Response** indicates that the registration procedure was successful.

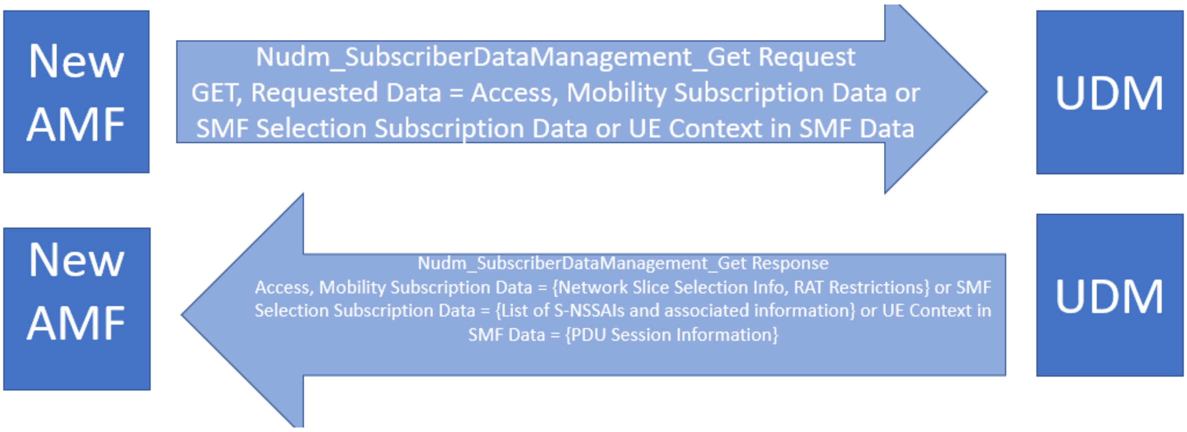

New AMF
Nudm_SubscriberDataManagement_Get Request
GET, Requested Data = Access, Mobility Subscription Data or SMF Selection Subscription Data or UE Context in SMF Data
UDM
New AMF
Nudm_SubscriberDataManagement_Get Response
Access, Mobility Subscription Data = {Network Slice Selection Info, RAT Restrictions} or SMF Selection Subscription Data = {List of S-NSSAIs and associated information} or UE Context in SMF Data = {PDU Session Information}
UDM

Figure 15: Subscriber Data Management Get Service

The UDM pushes out the Access and Mobility subscription data as well as the SMF selection subscription data via **Nudm_SubscriberDataManagement_Get** service (Requested data field indicator is set in accordance with the requested data) to this new AMF.

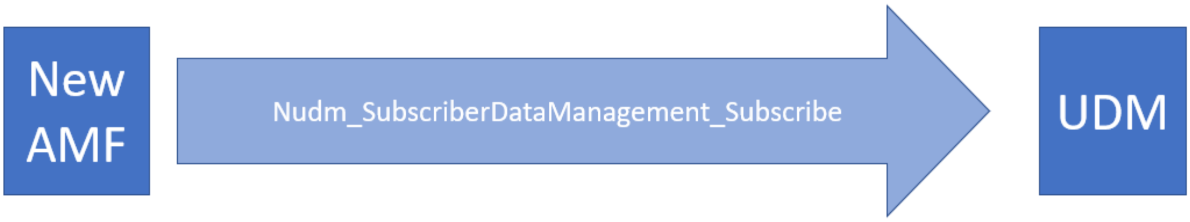


Figure 16: Subscriber Data Management Subscribe Service

Then the new AMF subscribes to the UDM regarding all the subscription data changes that may happen for this subscriber related to the subscriber data management via the **Nudm_SubscriberDataManagement_Subscribe** service.

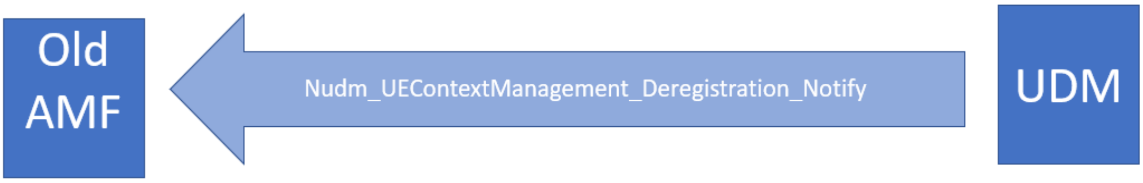


Figure 17: UE Context Management Deregistration Notify Service

At that point the UDM needs to send a **Nudm_UEContextManagement_Deregistration_Notify** which is a POST request sent to the callback Reference (deregCallbackUri field in Amf3GppAccessRegistration) as provided by the Old AMF during the registration. This procedure is initiated because the old AMF is no longer serving the subscriber and it has been deregistered in the UDM.

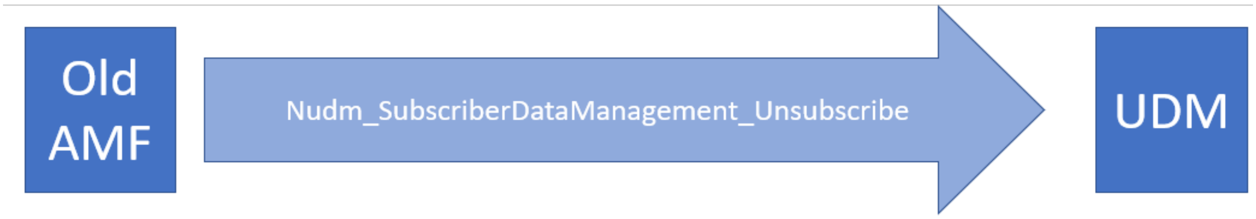


Figure 18: Subscriber Data Management Unsubscribe Service

The old AMF acknowledges via the **Nudm_SubscriberDataManagement_Unsubscribe** service. The new AMF then creates an MM context for the UE after getting the AMF related subscription data from the UDM. All the information needed to be shared across the home PLMN or across any visited PLMN depending on the SLA agreement is stored in the UDM. Among other critical information the subscribed NSSAI or RAT restrictions are stored in the Access and Mobility Subscription data of the UDM so that every AMF in HPLMN or VPLMN will be able to access to that information. Furthermore, in the SMF Selection Subscription Data, the Subscribed NSSAI Information is stored, which includes a list of S-NSSAIs and associated information (DNN Info) for establishing a PDU Session to every associated DNN. Overall, this mechanism reduces dramatically the signaling would be needed to be exchanged, as providing the needed information to every node in a direct way, without the need for communication with the last serving Entities.

4.2.11 Update New AMF Policy Association with the PCF

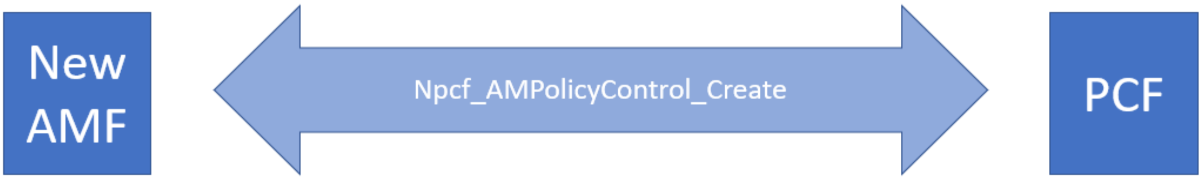


Figure 19: AM Policy Control Create Service

The serving AMF fetches the policies from the selected PCF based on the SUPI and creates the required associations using the **AMPolicyControl_Create** service.

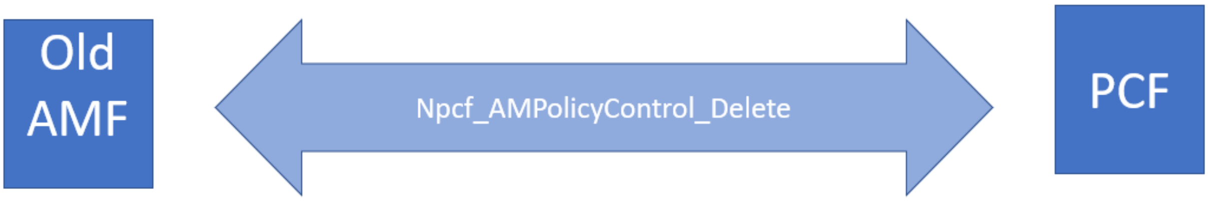


Figure 20: AM Policy Control Delete Service

The serving AMF sends a request, gets a response and thus retrieves the operator's defined UE Policy and/or Access and Mobility Control Policy. If the old AMF previously requested UE context to be established in the PCF, the old AMF transmits a **Npcf_AMPolicyControl_Delete Request** to terminate the UE context in the PCF. The PCF responds by transmitting a **Npcf_AMPolicyControl_Delete Response** message to the old AMF signaling the successful delete with the "204 No Content" HTTP response code. If the Registration type indicated by the UE is a periodic registration update, then this procedure is omitted. In case the access and mobility policies are deployed AMF initiates establishment of the AMF policy association with the PCF to get all the policies for this subscriber and then every AMF is able to get those policies from the UDM when it was getting all the subscription data. So, an association with the PCF is created for this given subscriber. This is an optional procedure taking place only when PCF is deployed.

4.2.12 SMF Discovery and Selection

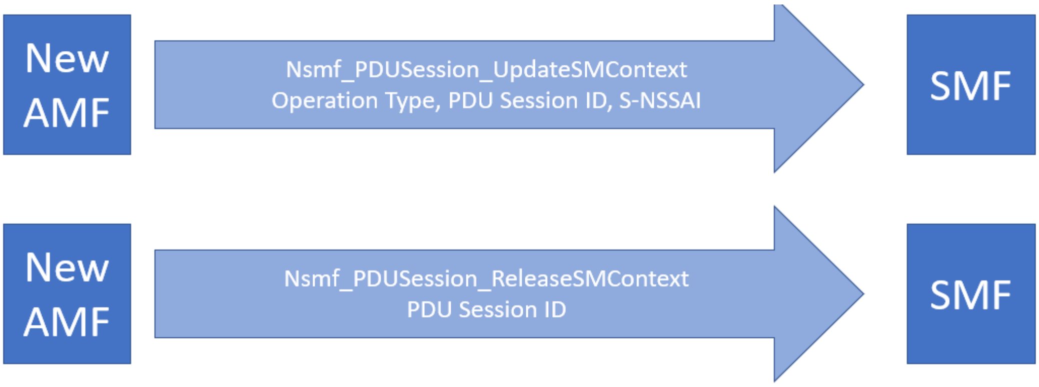

New AMF
Nsmf_PDUSession_UpdateSMContext
Operation Type, PDU Session ID, S-NSSAI
SMF
New AMF
Nsmf_PDUSession_ReleaseSMContext
PDU Session ID
SMF

Figure 21: PDU Session Update/Release SM Context Services

Generally, the SMF discovery and selection within the selected Network Slice Instance (NSI) is initiated by the AMF when a Session Management message to establish a PDU Session is received from the UE. When establishing a PDU session associated to an S-NSSAI and a DNN, a UE that is registered in a PLMN and has obtained an Allowed NSSAI, shall indicate in the PDU Session Establishment procedure the S-NSSAI and, if available, the DNN that the PDU Session is related to. The NRF is used to assist the discovery and selection tasks of the required network functions for the selected NSI. The AMF queries the NRF to select an SMF in a NSI based on S-NSSAI, DNN and other information e.g., UE subscription and local operator policies, when the UE triggers the establishment of a PDU Session. The selected SMF establishes a PDU Session based on S-NSSAI and DNN. The same procedure is initiated in a mobility or periodic registration updating scenarios and especially when an AMF relocation is performed. In such case the new AMF notifies each of the associated SMFs with the last serving AMF. The new AMF verifies PDU session status (if it is present) from the UE with the available SMF information retrieved from the old AMF. The AMF selects the SMFs based on the S-NSSAI included in either the C-NSSAI or the A-NSSAI and the UE's default configuration information included in the subscription information in the UDM. There may be fewer number of SMFs selected than the number of S-NSSAIs, which depends on the network slice configuration. The AMF may query the Network Repository Function (NRF) if it is not able to determine the appropriate SMFs (if they are not preconfigured from the Operator). In case of mobility registration updating or periodic registration updates the UE may forward an indication via PDU Session status for active user plane connections. Subsequently with the initial registration procedure a PDU Session Establishment Procedure shall be initiated in order for the UE to establish one or more user plane tunnels depending on the number of allowed S-NSSAIs. The

PDU session establishment request may be generated by the UE and included in the NAS request, or it may be generated by the default or serving AMF based on the parameters provided by the UE in the NAS request. The AMF forwards the PDU Session Establishment Request to the selected SMF(s). The SMF(s) may obtain additional PDU session information from the UDM (e.g., DNN, Session and Service Continuity (SSC) mode, the PDU session related parameters include the SST/SD or SMF ID) to trigger a PDU Session Establishment procedure. So, the AMF uses the requested SST/SD to select an appropriate SMF and sends the PDU session establishment request to the selected SMF to establish a PDU session. Typically, the PDU Session Establishment Procedure to the network can either be established due to mobile originated signaling or it can be mobile terminated. If it is mobile originated data, the UE identifies in the registration request all the PDU Session IDs for which uplink user plane traffic is pending and that is an indication for the AMF to reach out to the SMF via an **Nsmf_PDUSession_UpdateSMContext** informing that the Subscriber is going to actively use the following PDU Session IDs and then the SMF can interact with the UPF the User Plane Function and can make sure that those PDU Sessions are active for the UE to use. Although in a scenario where certain PDU Sessions are not available in the SMF. In this case the SMF is constant but under certain circumstances the SMF is also changing. In case a new SMF is probably not able to serve all the PDU Session IDs that this user has requested, the AMF actually reaches out to the old SMF via **Nsmf_PDUSession_UpdateSMContext** informing to retain active all the session management contexts for the PDU Session IDs that can't be served by the new SMF and release all the others via **Nsmf_PDUSession_ReleaseSMContext.** In the scenario that the SMF is constant, the new AMF reaches the SMF via a **Nsmf_PDUSession_UpdateSMContext** to inform about the newly associated PDU Sessions and a **Nsmf_PDUSession_ReleaseSMContext** to release the specified PDU Sessions that in no longer associated.

4.2.13 Registration Accept

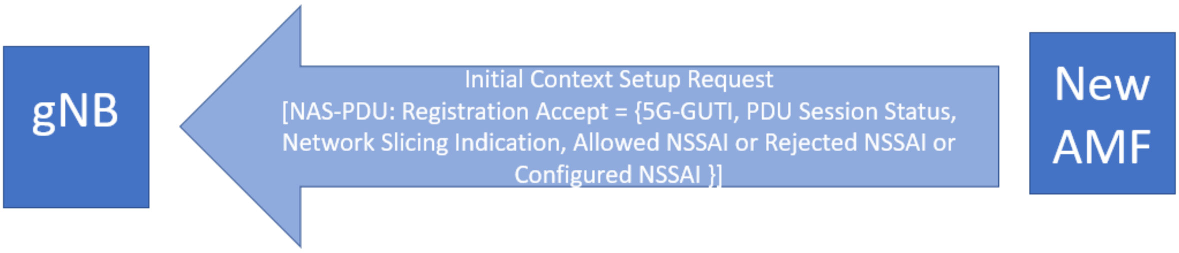

gNB
Initial Context Setup Request
[NAS-PDU: Registration Accept = {5G-GUTI, PDU Session Status, Network Slicing Indication, Allowed NSSAI or Rejected NSSAI or Configured NSSAI }]
New AMF

Figure 22: Initial Context Setup Request/ NAS PDU Registration Accept

After the SMF Selection and in case that the outcome of all the previously initiated procedures is successful, the AMF sends a NAS registration accept. The Registration Accept message among other IEs includes the 5G-GUTI, PDU session status, NSSAI. The registration accept is indicating that the registration has been accepted. 5G-GUTI is included if the AMF allocates a new one. The AMF indicates the new SMF(s) that were selected for the specified S-NSSAI. The AMF indicates the PDU session status to the UE. The UE removes any internal resources related to PDU sessions that are not marked active in the received PDU session status. If the PDU session status information was in the Registration Request, the AMF shall indicate the PDU session status to the UE. The NSSAI includes the allowed S-NSSAIs. The allowed S-NSSAIs includes all of the services that are available for the UE. If the UE de-registers from the network and needs to re-register, the UE sends a registration request including the 5G TMSI and the s_nssai_list in RRC and/or the A-NSSAI in NAS. In an unsuccessful scenario the UE might also obtain one or more rejected S-NSSAIs with cause and validity of rejection from the AMF carried by a Registration Reject message. An S-NSSAI may be rejected for the entire PLMN or for the current Registration Area. In both cases the UE shall not re-attempt to register to an S-NSSAI for the entire PLMN until is deleted by the network and shall not re-attempt to register to an S-NSSAI rejected in the current Registration Area until it moves out of the current Registration Area.

4.2.14 NR RRC Reconfiguration

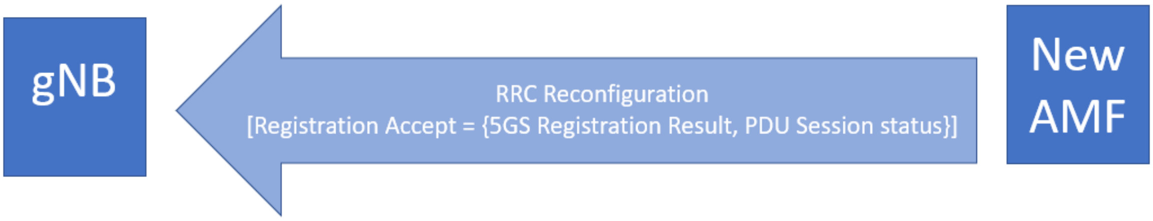

gNB
RRC Reconfiguration
[Registration Accept = {5GS Registration Result, PDU Session status}]
New AMF

Figure 23: RRC Reconfiguration

The receipt of the Registration Accept may trigger the RRC Connection Reconfiguration Procedure between the UE and the RAN for setting up radio bearers, setup a secondary cell and initiate UE measurements for the UE. The NSSAI along with the TMSI may be provided to the RAN by the new AMF.

4.2.15 Registration Complete

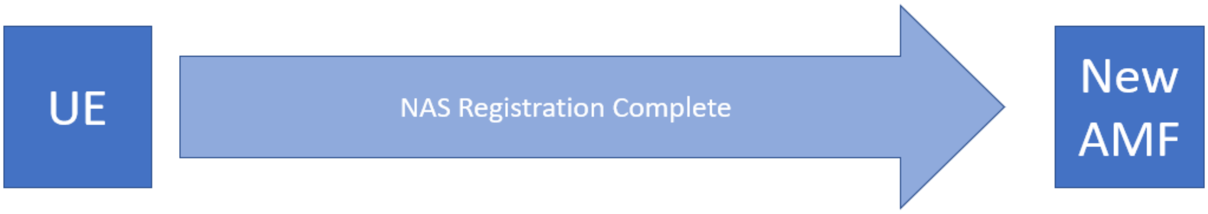


Figure 24: NAS Registration Complete

The UE sends a Registration Complete message to the AMF to acknowledge if a new Temporary User ID was assigned.

4.2.16 Update UE Policy Association with the PCF

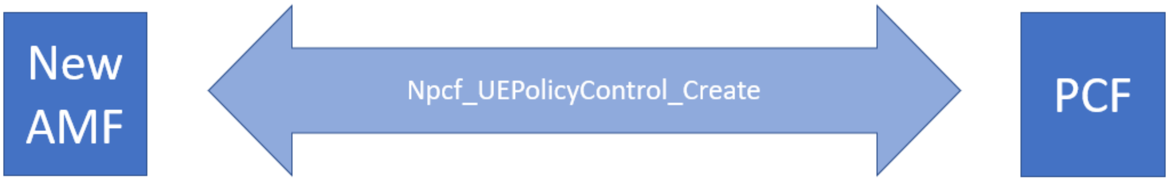


Figure 25: UE Policy Control Create Service

In case UE policies are deployed the AMF initiates establishment of UE policy association with the PCF. In 5G the PCF is able to provision UE policies directly into the UE. The PCF creates an association with the AMF via the **Npcf_UEPolicyControl_Create** message and then those policies can be pushed on to the UE by the NAS protocol (N1 Interface) from the AMF to the UE.

4.3 Security Infrastructure in 5G Architecture

4.3.1 Aspects in the Security Infrastructure

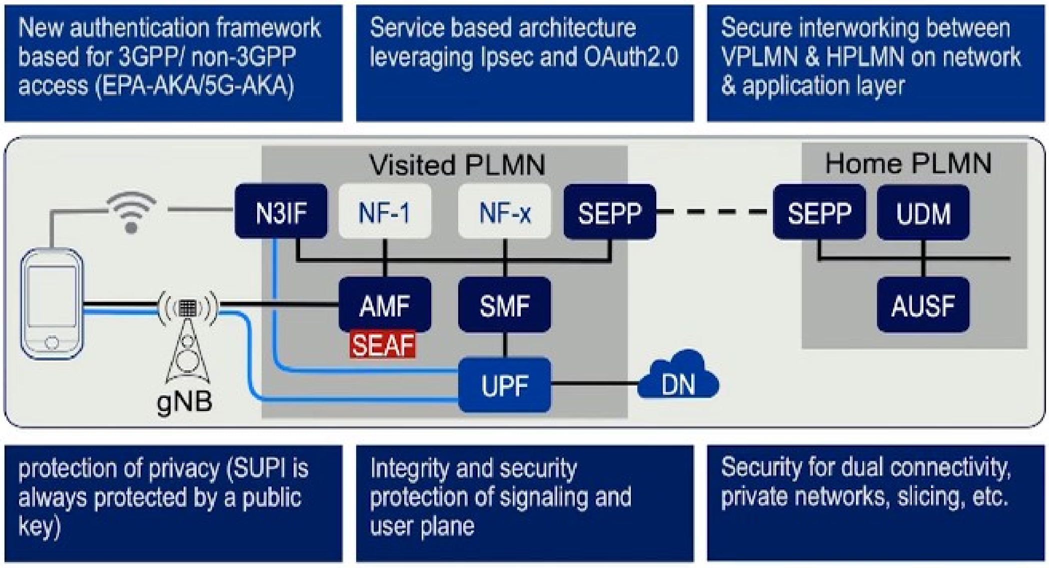

New authentication framework based for 3GPP/ non-3GPP access (EPA-AKA/5G-AKA)
Service based architecture leveraging Ipsec and OAuth2.0
Secure interworking between VPLMN & HPLMN on network & application layer
Visited PLMN
Home PLMN
N3IF
NF-1
NF-x
SEPP
SEPP
UDM
AMF
SEAF
SMF
AUSF
UPF
DN
gNB
protection of privacy (SUPI is always protected by a public key)
Integrity and security protection of signaling and user plane
Security for dual connectivity, private networks, slicing, etc.

Figure 26: Main Security Aspects Covered by 3GPP Rel.15.

5G is a decentralized network with many components deployed and lots of network functions. Several security levels have been introduced ensuring for the privacy and the integrity of the information between these functions in the Packet Core as well as between the nodes in the RAN [19]. The network entities in the SBA should communicate in an environment where their Signaling would be transmitted through a secure encrypted interface. Also, 5G introduces a kind of communication between the serving Network and the home network so both have to establish a trust model, so they need to verify each other first. Also, another level of privacy protection is introduced by keeping the permanent identity of a subscriber concealed as long as possible and an enhancement based on the user and control plane data as should be both encrypted and integrity protected.

4.3.2 Trust Model in 5G

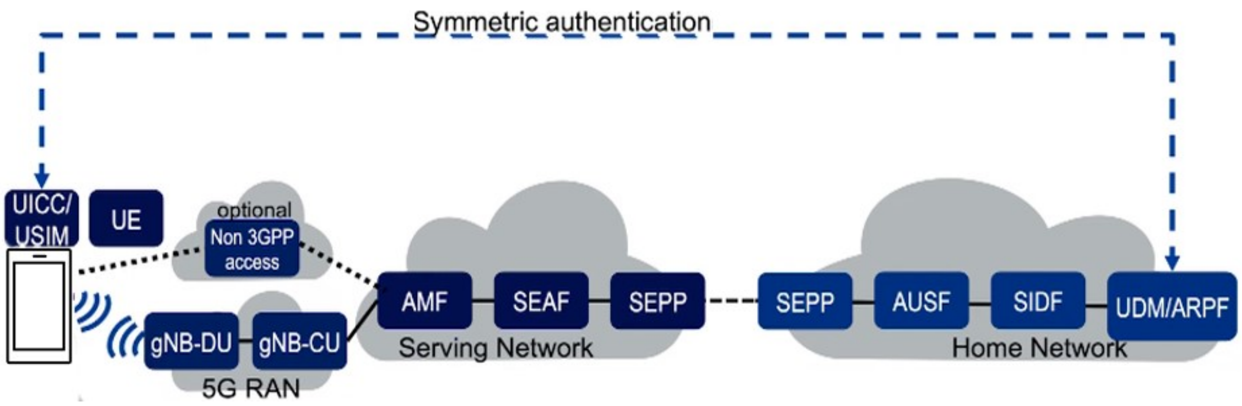


Figure 27: Trust Model in 5G

Trust is of crucial importance for 5G. The trust model is constructed by having secret master key and then the authentication procedure is instantiated as a symmetric comparison. This master key is stored in the SIM card or UICC and the other part of this master key is stored in a function in the home network, in the Authentication Credential Repository Function (ARPF). Then a derivation process is following. There are some entities in the home network like an authentication function. Authentication function is responsible for the initial authentication. In the other part the Serving's Network Security Anchor Function (SEAF) communicates with the home network and this communication is running via the Security Entity Proxy (SEPP) resulting to an encrypted communication between the two networks. Also, 5G allows the communication not only over 5G radio network but also over Wireless Lan and non 3GPP technologies.

4.3.3 5G Initial Authentication Aspects

In the authentication mechanism [18][20] a new scheme has been introduced known as 5G authentication and key agreement (AKA) procedure. The master key in the authentication vector not only verifies in terms of authentication by the home network but also by the serving Network. So, there is a kind of Two Step authentication procedure with a hash function as the first part, but the major part is that the home network decides about the successful authentication to a Serving Network and afterwards the Serving network will get that feedback from the home network.

4.3.3.1 Subscriber Protection

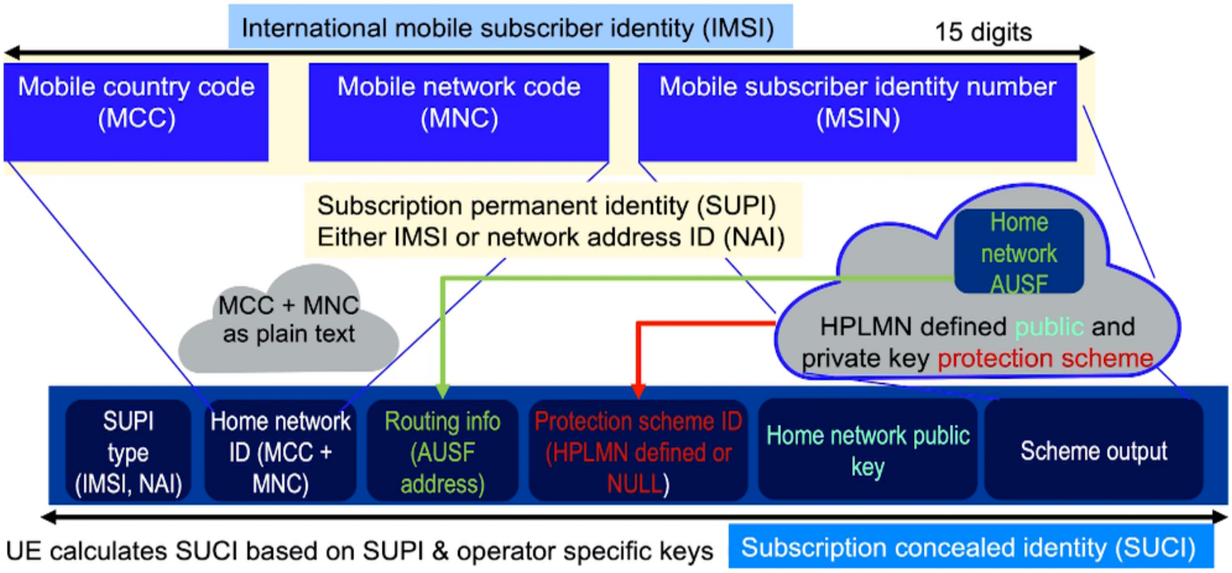

International mobile subscriber identity (IMSI)
15 digits
Mobile country code (MCC)
Mobile network code (MNC)
Mobile subscriber identity number (MSIN)
Subscription permanent identity (SUPI)
Either IMSI or network address ID (NAI)
MCC + MNC as plain text
Home network AUSF
HPLMN defined public and private key protection scheme
SUPI type (IMSI, NAI)
Home network ID (MCC + MNC)
Routing info (AUSF address)
Protection scheme ID (HPLMN defined or NULL)
Home network public key
Scheme output
UE calculates SUCI based on SUPI & operator specific keys
Subscription concealed identity (SUCI)

Figure 28: Privacy Model in 5G: SUPI and SUCI

Further to the protection of the subscriber’s permanent identities like International mobile subscriber Identity (IMSI) will also adopted in 5G and it is called as Subscriber Unique Permanent Identifier (SUPI). Also, there is an alternative to IMSI which is the network address identifier (NAI) which is also a globally unique permanent identity. By do not disclose this permanent identity in clear and plain text, 3GPP introduces a new concept of concealed value the subscription concealed identity (SUCI). So SUPI is being concealed by using a ciphering algorithm with predefined private and public keys by constructing SUCI such that only the home network operator will have this public key to perform the deciphering. So, the UE will contact the serving network by using its self-constructing SUCI and the serving Network will contact the home operator and only at this stage the SUCI will be deciphered into the real identity. This real identity or SUPI is disclosed to the Serving Network only after successful authentication. So, attacks like network spoofing whereas intruders try to emulate the network, is avoided with this mechanism where the real identity of the subscribers is deconcealed after a successful authentication.

4.3.3.2 A Drawback from the Continuous Key Derivation

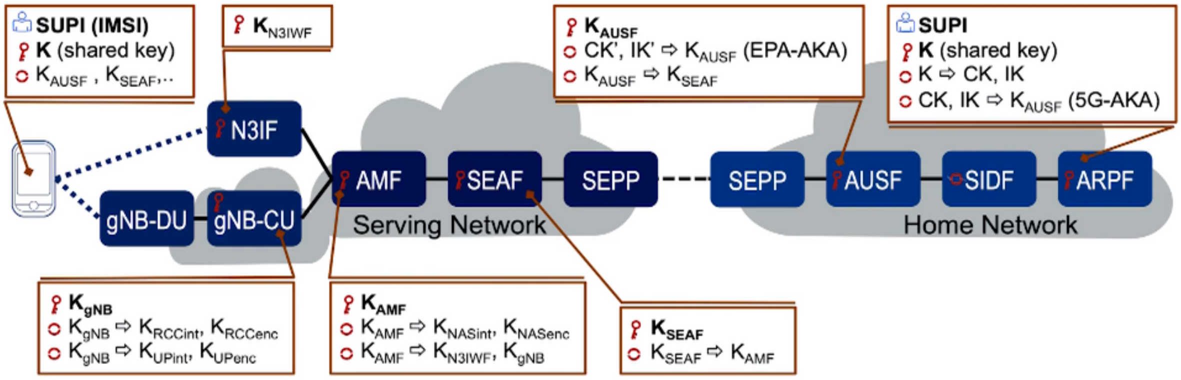

SUPI (IMSI)
K (shared key)
K_{AUSF} , K_{SEAF},..
K_{N3IWF}
K_{AUSF}
CK', IK' ⇨ K_{AUSF} (EPA-AKA)
K_{AUSF} ⇨ K_{SEAF}
SUPI
K (shared key)
K ⇨ CK, IK
CK, IK ⇨ K_{AUSF} (5G-AKA)
N3IF
AMF
SEAF
SEPP
SEPP
AUSF
SIDF
ARPF
gNB-DU
gNB-CU
Serving Network
Home Network
K_{gNB}
K_{gNB} ⇨ K_{RCCint}, K_{RCCenc}
K_{gNB} ⇨ K_{UPint}, K_{UPenc}
K_{AMF}
K_{AMF} ⇨ K_{NASint}, K_{NASenc}
K_{AMF} ⇨ K_{N3IWF}, K_{gNB}
K_{SEAF}
K_{SEAF} ⇨ K_{AMF}

Figure 29: Key Generation Hierarchy in 5G

In scenarios where the identification procedure is initiated and in case that the serving network asks the UE regarding it's SUCI, the provided SUCI will always be different as it is continuously recalculated. This key derivation process has a drawback as the home network operator is involved continuously when serving operator needs to validate a subscriber's SUCI. So, the Security procedure will be long lasting and inherently the latency will be increased. Therefore, 3GPP introduces an enhancement whereby using the master key in the home network, the authentication server will calculate out of this master key the so-called Anchor Key K_SEAF and will be stored in the serving network. Based on this anchor key all the following keys for example for radio resource integrity for user plane ciphering and for a handover to another technology always use a key derived out of this anchor Key and this allows a very fast and flexible security massive mechanism but still on a certain security level.

4.3.4 Security Aspects in Network Slicing

The majority of 5G Operators are implementing their Packet Core to be Virtualized End to End. In Network Slicing which comes inherently with 5G the network is being isolated and split into multiple instances assigned to different use cases. In a scenario where an unauthorized intruder, attacks in a particular slice to intrude into the virtualized Packet Core of that Slice, the other existing slice instances remain unaffected as they are isolated. Although is an operator's decision of the architecture and the orchestration tools will be used for the Packet Core, should have isolated the devices needed for data traffic and critical session management traffic or even access management traffic from the common dedicated devices per slice. Therefore, an attack on these elements does not affect the signaling in the other slices. The Packet Core Functions where a lot of Signaling traffic is passed through them, should be isolated in different containers or different virtual machines.

4.3.5 Security Mechanisms for Packet Core

The communication between the network functions especially on the service-based architecture interfaces if they are 3GPP compliant is encrypted and integrity protected. For their communication is used TLS 1.2/1.3 and OAUTH 2.0 so before establishing the communication channel they mutually authenticate between each other, and they use tokens to communicate between each other. According to [1] a trust model has been designed around critical elements in the network like UDM or ARPF where more security levels are applied, and thus more control mechanisms designed around them in a way to protect them not just with one layer but with more layers like network or kernel isolation or with firewalls in designated places where those controls are needed. From a physical infrastructure perspective depending on the requirements and the security mechanism that an operator would like to adopt to its Packet Core, more controls can be added as going out of the layers and thus by crossing from one trust level to another.

5. THE ETSI PERSPECTIVE OF NETWORK SLICING

5.1 Virtualization of the Physical Network Infrastructure Resources

Network slicing enables the definition of multiple isolated virtual networks on a single physical networking infrastructure. Each network slice is created to meet the specific requirements of a set of applications and services. Additionally, a network slice is viewed as a logical end-to-end network that can be created dynamically enabling multiple services with diverse requirements to share available networking infrastructures. The network slicing consists of one, isolated subset of the available virtual resources like computation, networking, storage and secondly, a set of rules for identifying the traffic that will run on those slices. A network slice will consist of a set of virtual resources and the traffic flows associated with it. The physical resources able to be virtualized[21] and thus sliced are the following:

5.1.1 Bandwidth

Each slice and especially in the same Tracking area should have its fraction of bandwidth on a link in a dedicated way.

5.1.2 Topology of a Network

Each slice should have its own view of the network nodes. The nodes can be switches, routers and generally network communications devices. Meantime the connectivity between these network nodes should be also available for different slices via offered underlying mechanisms in order for the isolation to be retained.

5.1.3 CPU Utilization

Each slice should be assigned to the computational resources that are cutting to the requirements of different applications or services.

5.1.4 Storage

Storage is very important for different network devices. Varying levels of storage capacity should be managed based on the needs and the traffic in every slice.

5.1.5 Control Plane

If there is a packet in and packet out conforming with the rules in the forwarding tables, these tables and other control plane resources should also be sliced.

5.1.6 Traffic

A specific portion of the traffic to one or more virtual networks should be associated with a slice to be cleanly isolated from the remaining underlying network. This is very important; it also refers to the isolations of the network slice.

5.2 The 3 Core Layers of the Network Slice Architecture

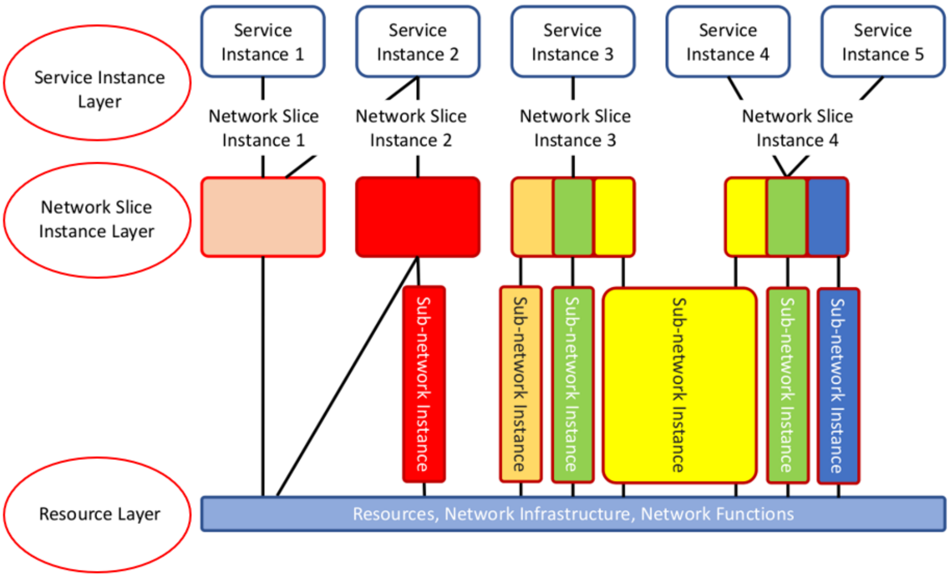


Figure 30: Network Slicing Core Layers

5.2.1 The Service Instance Layer

The service instance layer hosts the services of applications which are expected to be supported by the network. These services can either be provided by the network operator or by a third party.

5.2.2 The Network Slice Instance Layer

The network slice instance represents a collection of resources from the resource layer below, to form a network slice. The network slice instance provides the network characteristics required by a service instance. The network slice instance may be shared across multiple service instances, which are provided by a network operator. The network slice instance can consist of none, one or more Sub-network Instances, which may be shared by another network slice instance.

5.2.3 The Resource Layer

The resource layer hosts different sub-Network instances where each sub network instance represents a resource. A resource or a network function can serve one or more network slice instances. At this layer, network slice management function is performed by the resource orchestrator, which is composed of NFV Orchestrator (NFVO), and of application resource configurators.

5.3 Network Slicing Lifecycle Management and Monitoring

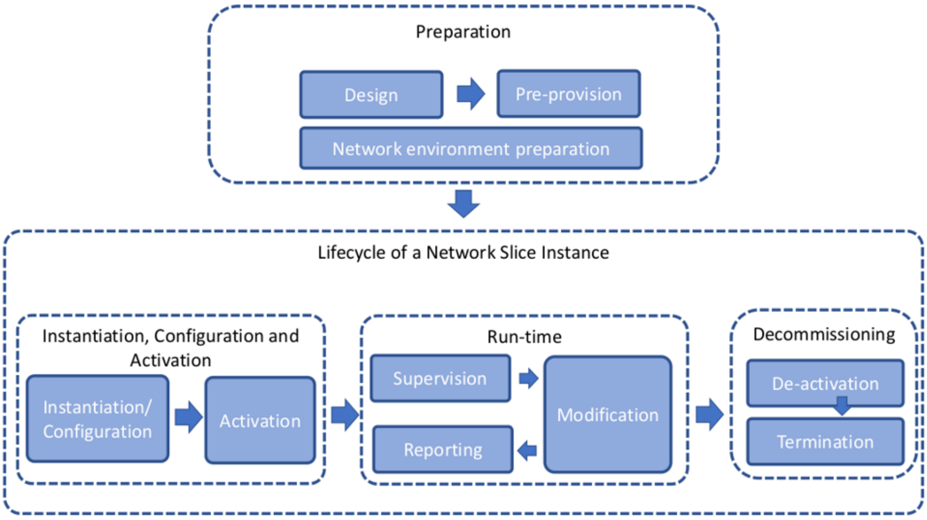

Preparation
Design
Pre-provision
Network environment preparation
Lifecycle of a Network Slice Instance
Instantiation, Configuration and Activation
Instantiation/ Configuration
Activation
Run-time
Supervision
Reporting
Modification
Decommissioning
De-activation
Termination

Figure 31: Network Slice Lifecycle

According to Open Network Foundation (ONF) "Network slicing is a mobile networking platform where elastic and scalable access and connectivity related capabilities are provided as a service to customers using 3GPP standardized Technologies." The elasticity and scalability are somewhat similar [22]. Elasticity gives the ability to grow or shrink the networking resources, the spectrum and the compute storage dynamically as needed. An important aspect is also that the customers are not necessarily just the end-users, they could be entire enterprises or internal customers for an operator serving specific purposes. The architecture of the network should be configurable in such a way, by defining some slice attributes to be able to provide a specific latency, a specific throughput, reliability, mobility, locality etc. Network slicing effectively aims to transform a mobile network to a network cloud with distributed physical and services resources. In that sense network slicing is the paradigm with which operators can monetize their distributed Network clouds. A network slice is created only with the necessary network functions and network resources at a given time. They are gathered from a complete set of resources and network virtual network functions and orchestrated for the particular services and purposes defined by the slice attributes. A network slice is a dynamic entity therefore its lifecycle has to be managed. The network slice lifecycle management is (creation, update, deletion) is managed by the network slice orchestrator. The slice orchestrator according to requests sent either by the orchestrator operator or from 3rd parties or even by the end-users creates a new slice instance that is based on a slice template that is stored in the slice template repository. However, it takes into account slice operator (owner) preferences (policies).

5.3.1 Preparation Phase

So, the process for the life cycle of network slicing starts with design and preparation of a slice template. Significant parameters should be taken into account at this stage like network slice capacity planning, on-boarding and evaluation of the network slice requirements and preparing the network environment in order to design the network slice template. A network slice template is a description of components, structure, and configuration of a slice. The slice itself does not exist in this phase, and it will be built from the template in the second phase.

5.3.2 Commissioning Phase

The commissioning phase is reflected on the instantiation request to create, configure and activate the slice. The resources and network functions are created, installed, and configured. The slice is built from the template (using specific instance information), installed, configured, and activated. The creation of a network slice instance can include creation and/or modification of the network slice instance constituents [23].

5.3.3 Operation Phase

The operation phase includes the activation, supervision, performance reporting (e.g., for KPI monitoring), resource capacity planning, modification, and de-activation of a network slice instance in order to meet quality of service requirements. Provisioning in the operation phase involves activation, modification and de-activation of a network slice instance.

5.3.4 Decommissioning Phase

The provisioning of a network slice instance in the decommissioning phase includes decommissioning of non-shared constituents if required and removing the network slice instance specific configuration from the shared constituents. After the decommissioning phase, the network slice instance is terminated and does not exist anymore.

This life-cycle process is handled by the network slice controller or network slice manager. The slice manager can be accessed by a north-bound standardized Application Program Interface (API). Depending on the scenario, the operator might allow different actions in the API: create or delete slices, different levels of configuration, report, and monitoring.

5.4 Management and Orchestration Planes Architecture Principles

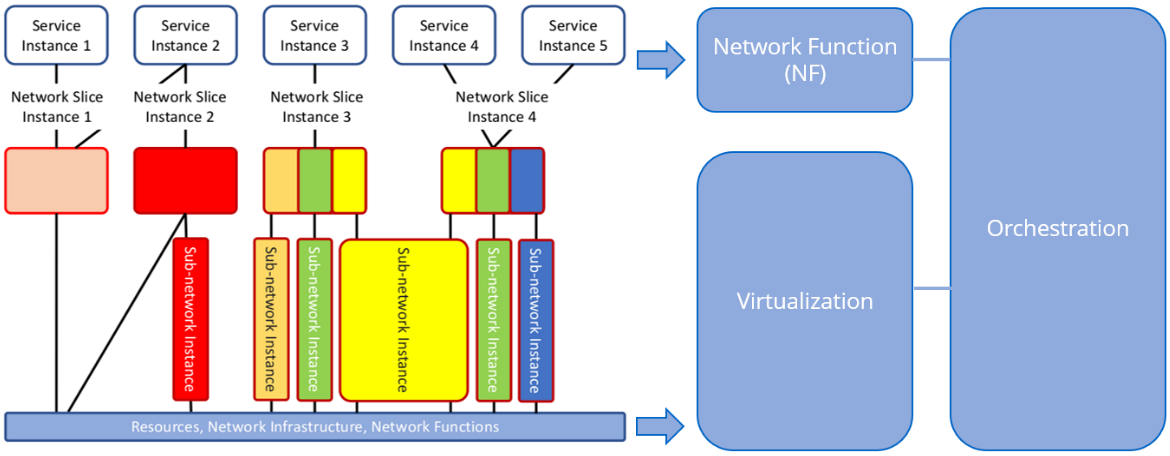


Figure 32: Network Slicing Layers Management and Orchestration

The basic idea of network slicing [24] is to slice the original network architecture. Beginning from a shared infrastructure or shared physical resources known as the resource layer. Slicing the resource layer in multiple logical and independent networks, sub-network instances are defined. The sub-network instances are configured to effectively meet different service requirements. These variated service requirements are forming the service instance layer. The service instance layer can be also a combination of different subnetwork instances. The underlying mechanism for the facilitation of the service instance layer is through a network function. A network function (NF) is a processing function in a network. It includes but is not limited to network nodes functionality, e.g., session management, mobility management, switching, routing functions, which has defined functional behavior and interfaces. Network functions can be implemented as a network node on a dedicated hardware or as a virtualized software function. Data, Control, Management, Orchestration planes functions are Network Functions. Thus, a Network function expresses elementary network functionalities that are used as built blocks to create a network slice. Moreover, the underlying mechanism in order to separate the shared physical resources and thus form different network functions is through virtualization. Virtualization provides an abstract representation of the physical resources under a unified and homogeneous schema, enabling a scalable slice deployment. Also, relying on network function virtualization NFV architecture mechanisms allowing the decoupling of each network function instance from the network hardware it runs on. So, after having the network functions and the virtualized network resources, a shared orchestration mechanism is needed for the interaction and communication between the entities. So, orchestration is the process that coordinates all the different network components that are involved in the life cycle of each network slice. In this context the Software-Defined Networking SDN Principle is deployed to enable a dynamic and flexible slice configuration. The main components of the network slicing architecture also comprise the SDN architecture principles.

5.5 Management and Orchestration Planes Architecture

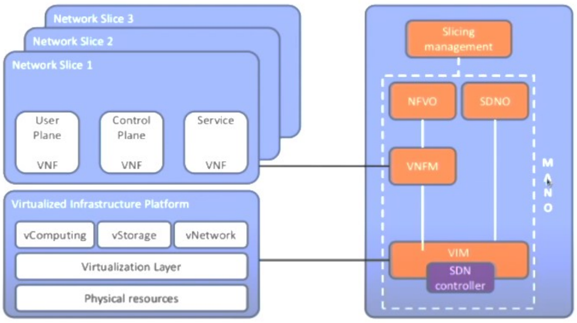


Figure 33: Management and Orchestration Planes Architecture

The corresponding architecture for the management [25] of network slicing consists of the following main components. The Virtualized Infrastructure Manager (VIM), the Network Slice Instance (NSI) and the Management and Orchestration (MANO). The Virtualized Infrastructure Platform provides virtual resources such as virtual computing, virtual storage, and virtual network to assign to one or more slices. Virtualization is performed through a virtual infrastructure manager (VIM). A Network slice instance is a collection of resources from the virtualized infrastructure platform. These resources are organized into different network slices according to the abstraction of characteristics of the services they facilitate. The Virtual Network Functions Manager VNFM is responsible for the Life Cycle Management (LCM), for the performance, fault and configuration management of the different network slices or VNFs based on the directives by the NFV Orchestrator. The Virtualized Infrastructure Manager (VIM), the Virtual Network Functions Manager (VNFM), the Network Function Virtualization Orchestrator (NFVO) and the SDN Orchestrator (SDNO) are a group of dedicated sub modules or functional blocks of the NFV-MANO, Management and Orchestration Architectural Framework. Each Virtual Infrastructure Manager includes one or more SDN controllers to enable virtualization [14]. The virtualization of the resources is performed by the virtualized infrastructure platform. Then the network slice instance can be instantiated in this architecture as a network function, orchestrated by the management and orchestration MANO component.

There are three typical architectures implementing the general architecture above [21].

5.5.1 Single Owner Single Controller Model

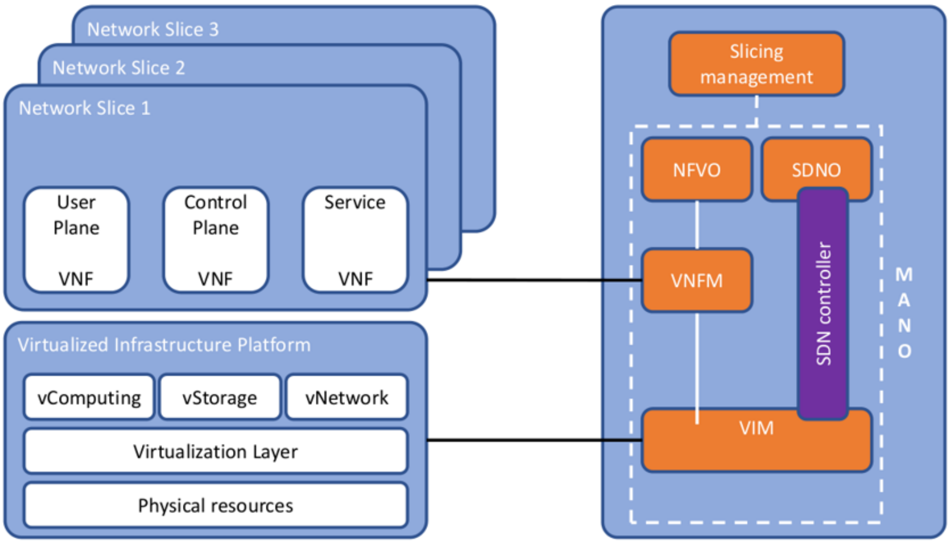


Figure 34: Single Owner Single Controller Model

Single Owner Single Controller is a network slicing architecture with a single controller or Orchestrator. The SDN mechanisms provide an abstraction of the network resources through the Northbound Interface of the SDN Controller. The management and orchestration functionalities are subsequently implemented on top of the SDN controller by exploiting the northbound interface. In this case the SDN controller, the main part of the MANO in the general architecture, operates as an SDN Orchestrator. This solution is ideal for small network operators especially in a case of a single infrastructure owner. In such a case one SDN controller will completely orchestrate all the different slices. Although this may represent a bottleneck in terms of performance and reliability as the presence of the single controller limits the programmability of the networking infrastructure especially in case multiple tenants desire deploying network services.

5.5.2 Single Owner, Multiple Tenants Model

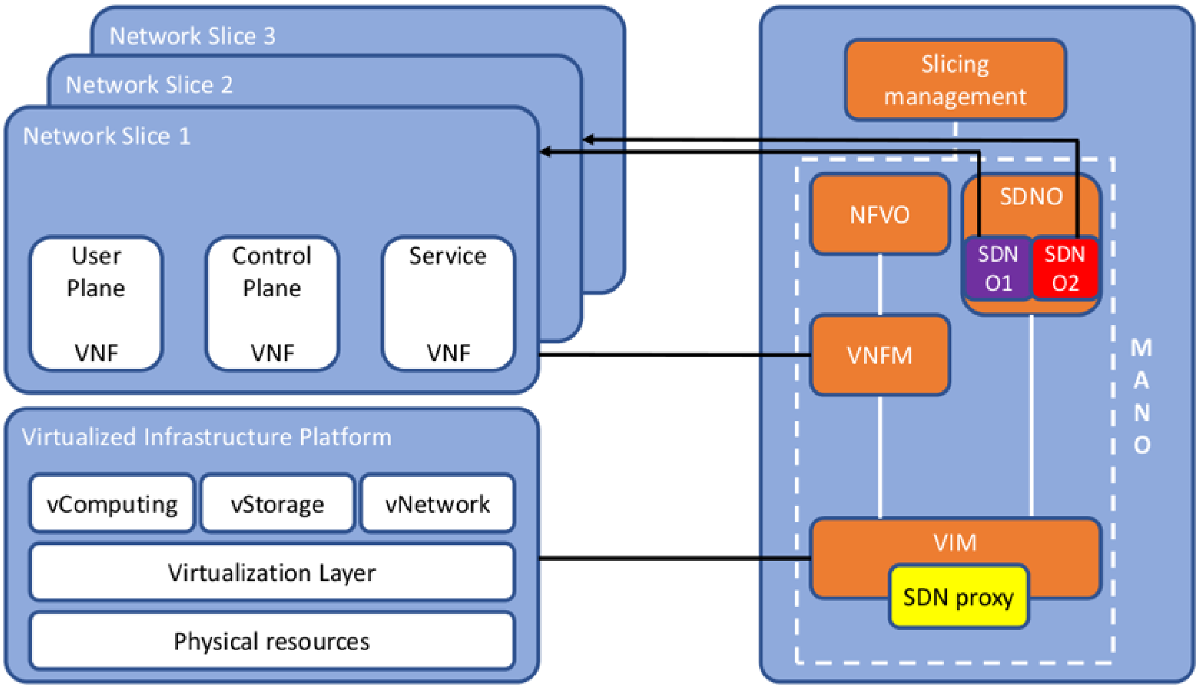


Figure 35: Single Owner, Multiple Tenants Model

The second architecture is called “Single Owner, Multiple Tenants”. The idea of supporting multiple virtual networks is now becoming common in many implementations from data centers to service provider networks. In this framework, the native technology to implement network slicing is by introducing an SDN Proxy typically controlled by the owner of the physical network infrastructure. The SDN Proxy provides an abstraction of the network forwarding path and thus gives the ability to slice the network. The SDN Proxy defines a hardware abstraction layer that is logically placed between controller and forwarding path in order to force rules and agreements defined in the network slice and to maintain isolations. This architecture is enabling multiple virtual tenants to deploy their own controllers or SDN Orchestrators on the shared infrastructure and meantime maintaining the isolation between different slice instances.

5.5.3 Multiple Owners or Multiple Tenants Model

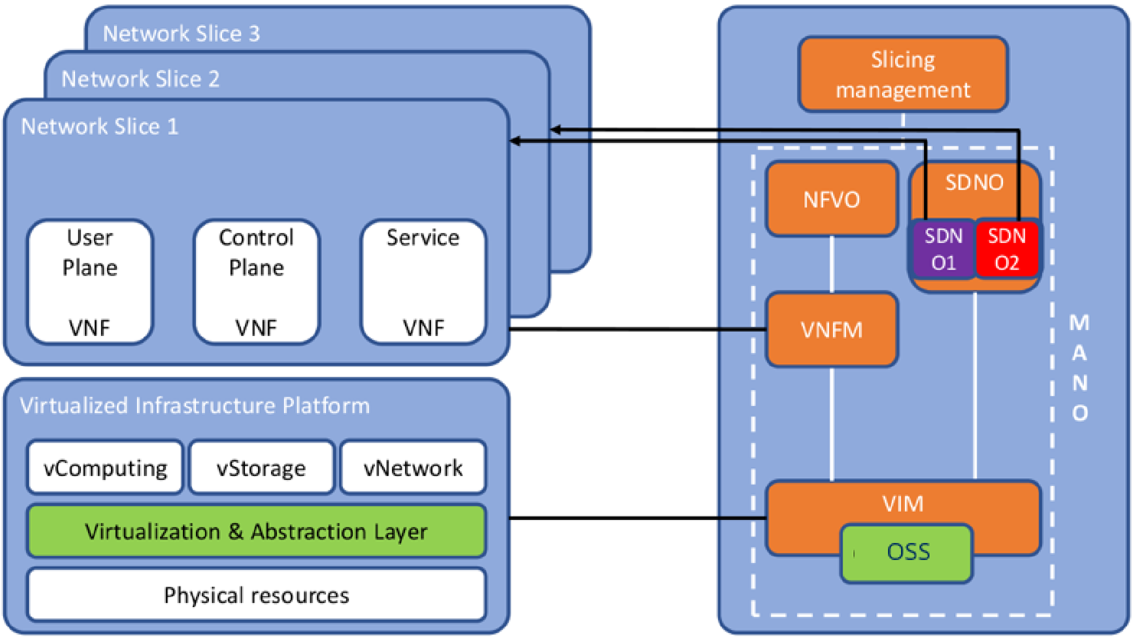


Figure 36: Multiple Owner or Multiple Tenants Model

The third architecture is called “Multiple Owners or Multiple Tenants”. For that architecture, the mechanisms offered via virtualization are essential to allow multiple tenants to specify their desired way for their resources to be connected. This implementation choice is expected to be independent from the service or infrastructure providers. The previous mentioned architectures are mainly focused on generating slices of the underlying network infrastructure by dividing or isolating the linked resources. Also, inherently with the generated slices one or more controllers are needed to handle a portion of the requirements needed to enable complete freedom for the tenants. For example, to give the ability to define their desired network topology and so on. Such architectures imply the need to introduce an advanced virtualization layer in the virtualized infrastructure platform and this virtualized or as called “Abstraction Layer” should be on top of the physical network but still under the virtualized resource. So, the virtualization and abstraction layer is located between the physical infrastructure resources and the virtual network controllers allowing to create isolated virtual networks with the topology specified by the tenants. Also, it allows to use any controller on the network operating system. The users will have the capability to use the whole address space and change the virtual network capabilities and functionalities on demand. The infrastructure owner(s) will retain control of its own virtual SDN network in a way that slices can be automatically recovered from physical failure.

6. THE GSMA PERSPECTIVE OF NETWORK SLICING

6.1 Different Services should be provided Concurrently.

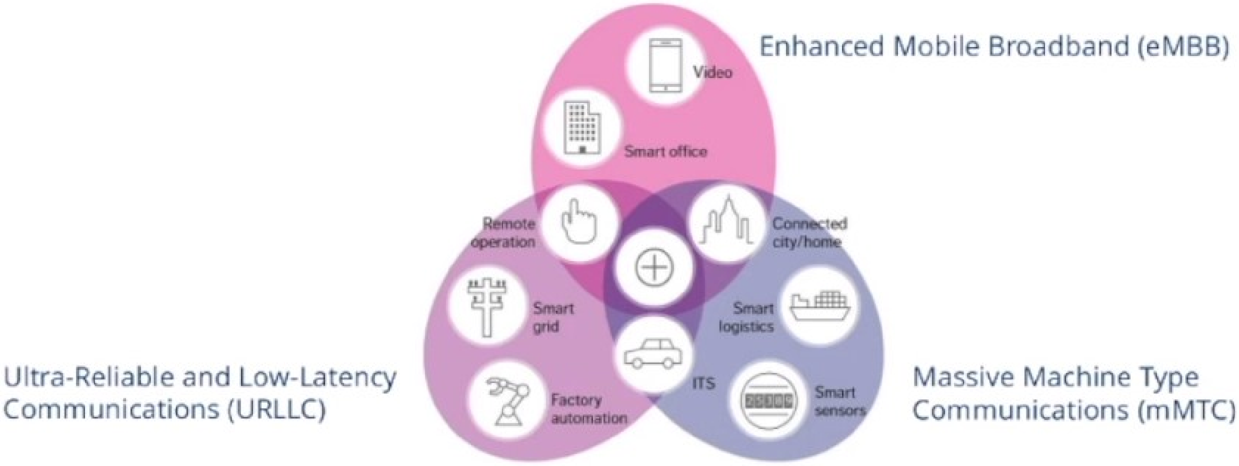

Enhanced Mobile Broadband (eMBB)
Video
Smart office
Remote operation
Connected city/home
Smart grid
Smart logistics
ITS
Factory automation
Smart sensors
Ultra-Reliable and Low-Latency Communications (URLLC)
Massive Machine Type Communications (mMTC)

Figure 37: 5G Triangle (Ericsson)

The whole 5G [26] will require one network with low delay, high throughput and high resilience. Although most services do not need low latency, high resilience and high throughput there are services in Non-Public Networks like the industries where the factory automation services need only low latency without requiring high throughput. On the other hand, video services demand high throughput but are more tolerant to delays. Also, services provided in the autonomous driving car need very low delay communications to ensure that the surrounding environment is safe and that the driving can be precise. In the meantime, the driver should be able to listen to the music or may be streaming some videos. For such additional services is also needed a very high throughput of the network. Once the car is close to its destination it wants to know if there are free parking slots in the parking area. In such situations several sensors in the parking area should be able to establish a connection with the car and thus inform if there is free space. For such services high resilience is required. From such an illustrative scenario can be derived, that in the future the networks need to serve customers or services with very different needs simultaneously.

6.2 Traditional Networks should Evolve.

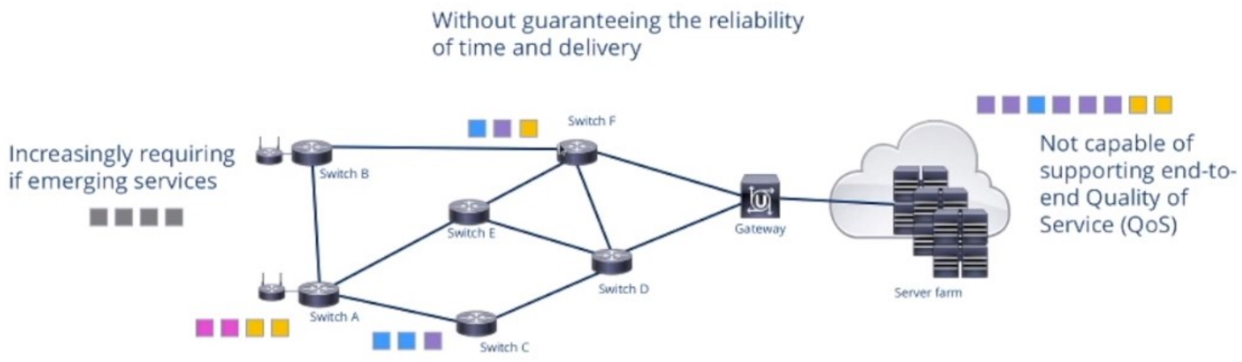

Without guaranteeing the reliability of time and delivery
Increasingly requiring if emerging services
Switch F
Switch B
Switch E
Switch D
Switch A
Switch C
Gateway
Server farm
Not capable of supporting end-to-end Quality of Service (QoS)

Figure 38: Best Effort Traditional Network

Today's networks are not able to fulfill this kind of demand. In a traditional network, several packets from multiple services are transmitted in the same network physical instance which basically offers a best-effort service meaning that each network device only tries their best to send out the packet. Therefore, the traditional networks resulting in any guarantees in delivery of data and secondly the traditional networks are commonly not capable of supporting quality of service especially end-to-end quality of services as the network devices do not have a holistic view of the whole network. In addition, taking into account the future scenario with the autonomous car, as the number of required services is expected to grow, the traditional network cannot support the increased requirements from these emergent services because of the best effort service it provides.

6.3 4G Architecture is only appropriate only for today's Specifications.

Also, in a single network architecture like 4G is not capable of supporting all requirements at the same time because it would need to balance among divergent solutions, and this could lead to several physical networks with individually determined quality of service parameter sets. Also, applications and services are continuously evolving, and it is extremely difficult to predict the future service requirements. So, there are several limitations in traditional networks and for future networks there is an emerging need to support two different types of mechanisms to provide end to end quality of service. The first mechanism is to enable rapid deployment of the related network configuration and the second mechanism should provide the capability for multiple quality of service parameter sets on one shared physical network infrastructure. This rational leads the idea of network slicing to be emerged.

7. QUALITATIVE ANALYSIS IN OPENSLICE AS A PROTOTYPE OSS DEPLOYMENT

7.1 OpenSlice Definition

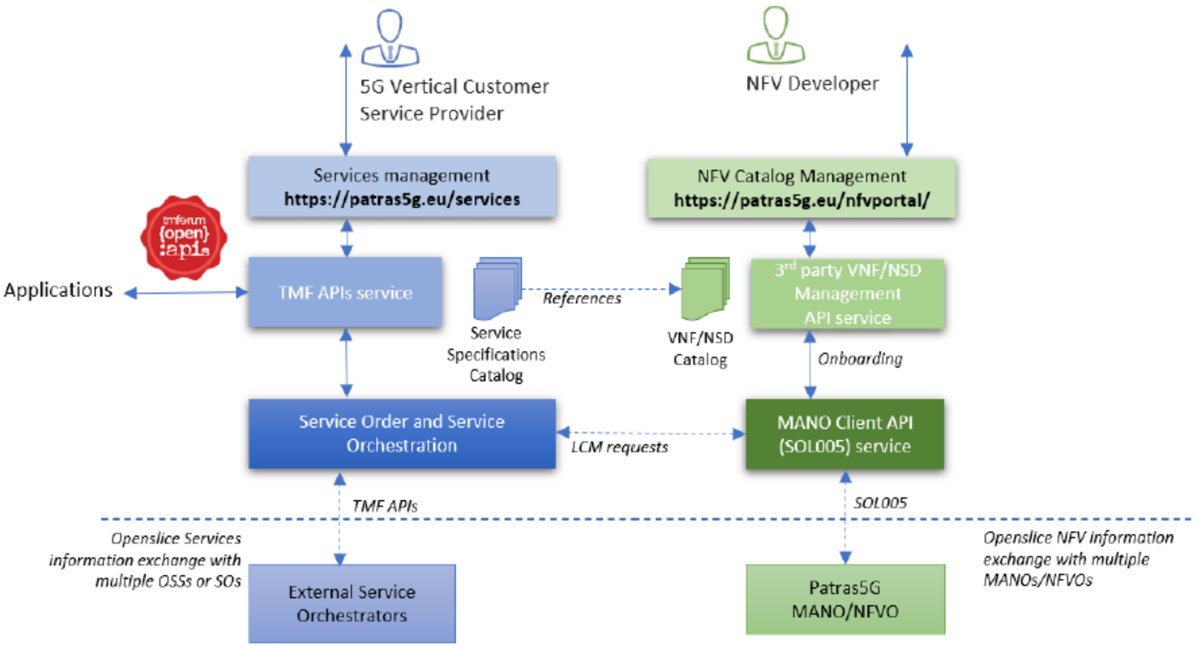

5G Vertical Customer
Service Provider
NFV Developer
Services management
https://patras5g.eu/services
NFV Catalog Management
https://patras5g.eu/nfvportal/
tmforum {open} :apis
Applications
TMF APIs service
References
Service Specifications Catalog
VNF/NSD Catalog
3rd party VNF/NSD Management API service
Onboarding
Service Order and Service Orchestration
LCM requests
MANO Client API (SOL005) service
TMF APIs
SOL005
Openslice Services information exchange with multiple OSSs or SOs
External Service Orchestrators
Patras5G MANO/NFVO
Openslice NFV information exchange with multiple MANOs/NFVOs

Figure 39: An Experimental Deployment of OpenSlice

OpenSlice [27] allows Vertical Customers to browse the available offered service specifications and allows NFV developers to onboard and manage VNF and Network Service artifacts. 3rd party applications can use OpenSlice through TMForum Open APIs. In general, OpenSlice offers the following main functionalities:

- Service Catalog Management: A Communication Services Provider will have the ability to manage the Service Catalog Items, their attributes, organize in categories and decide what to make available to Customers. Contains the representation of Service Specifications, either created from the provider defining service attributes, or by supporting the GSMA Generic Slice Templates (GST) as well as the VINNI Service Blueprint.
- Services Specifications: A CSP will be able to manage Service Specifications.
- Service Catalog Exposure: A CSP will be able to expose catalog to customers and related parties.

- Service Catalog to Service Catalog: OpenSlice able to consume and provide Service Catalog items to other catalogs.
- Service Order: The Customer will be able to place a Service Order.
- Service Inventory: The Customer and Provider will be able to view deployed Services status.

7.2 OpenSlice Architecture

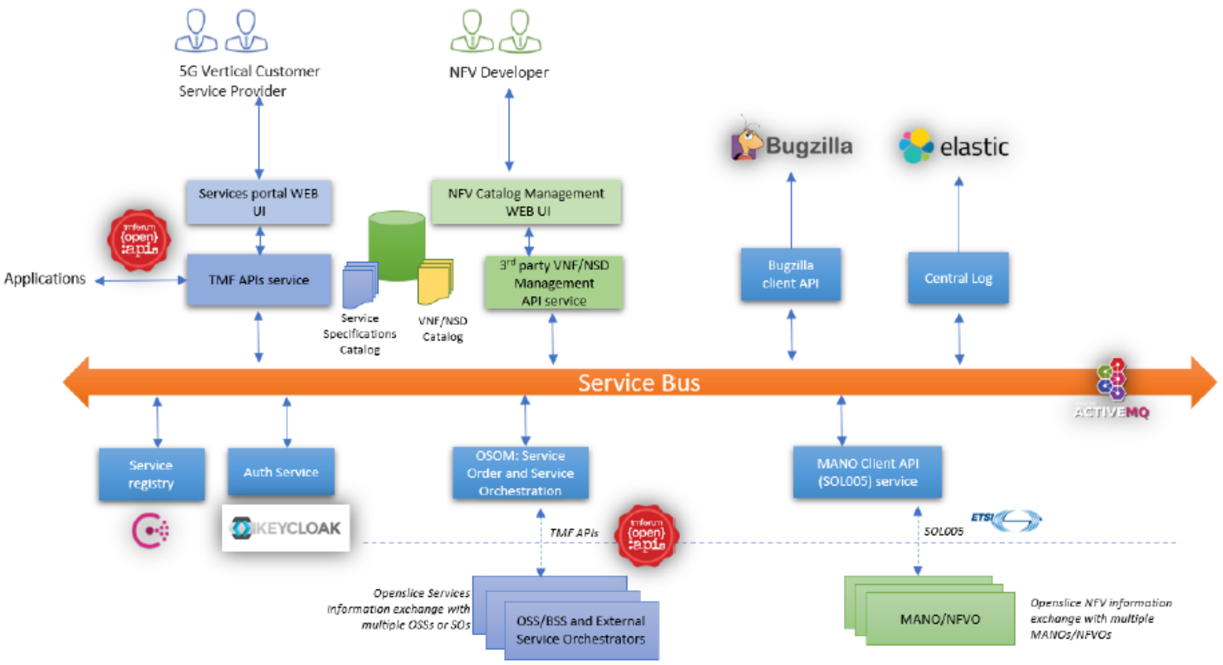

5G Vertical Customer
Service Provider
NFV Developer
Bugzilla
elastic
Services portal WEB UI
NFV Catalog Management WEB UI
Applications
TMF APIs service
Service Specifications Catalog
VNF/NSD Catalog
3rd party VNF/NSD Management API service
Bugzilla client API
Central Log
Service Bus
ACTIVEMQ
Service registry
Auth Service
KEYCLOAK
OSOM: Service Order and Service Orchestration
MANO Client API (SOL005) service
TMF APIs
SOL005
ETSI
Openslice Services information exchange with multiple OSSs or SOs
OSS/BSS and External Service Orchestrators
MANO/NFVO
Openslice NFV information exchange with multiple MANOs/NFVOs

Figure 40: OpenSlice Architecture

OpenSlice allows Vertical Customers browsing the available offered service specifications. Is made up of the following main functional components [28]:

Web frontend UIs that consist of mainly two portals:

- A NFV portal allowing users self-service management and onboarding VNFDs/NSDs to facility's NFVO.
- A Services Portal, which allows users to browse the Service Catalog, Service Blueprints specifications and the Service Inventory.
- An API gateway that proxies the internal APIs and used by the web front end as well as any other 3rd party service.

- A Message Bus where all microservices use it to exchange messages either via message queues or via publish/subscribe topics.
- An authentication server implementing Oauth2 authentication scheme.
- A microservice offering TMF compliant API services (e.g., Service Catalog API, Service Ordering API etc.)
- A microservice offering NFV API services (e.g., VNF/NSD onboarding etc.) and allows to store VNFDs and NSDs in a catalog
- A microservice that is capable to interface to an issue management system. For example, it raises an issue to all related stakeholders that a new Service Order is requested
- Central logging microservice that is capable to log all distributed actions in to an Elasticsearch cluster.
- A Service Orchestrator solution that will propagate Service Ordering requests to the equivalent SOs and NFVOs.

7.2.1 OpenSlice Portal

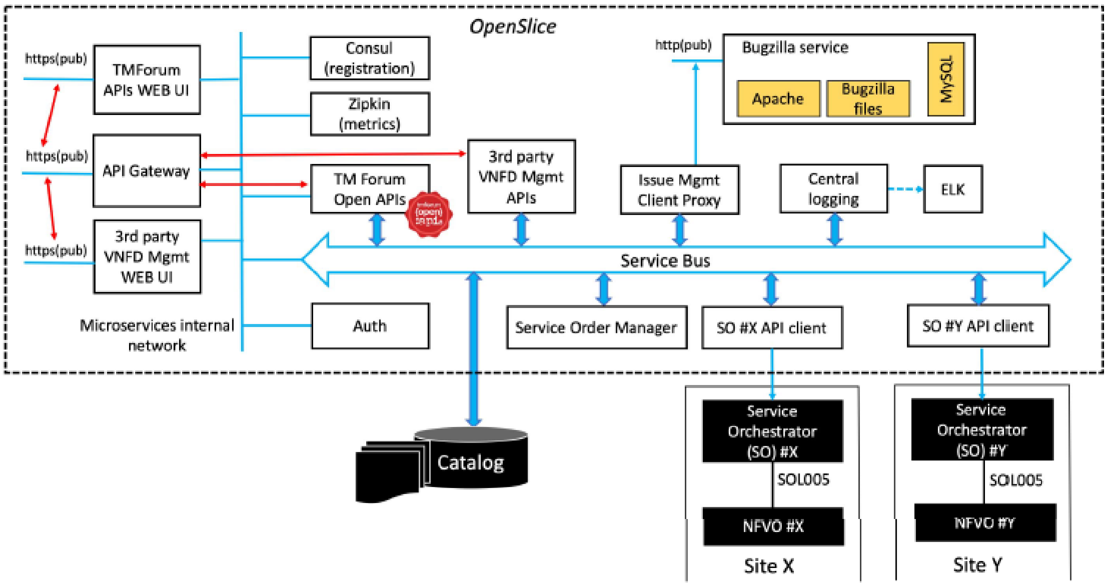


Figure 41: OpenSlice Front-End Back-End Interconnection

The portal's goal [29] is to expose the Customer Service Provider facility to verticals as a single network infrastructure rather than an interconnection of individual administrative domains, while concealing the mechanisms, protocols, and technologies used in each site to accomplish this. For the purposes of application testing and KPI validation, the portal is used to help verticals gain access to the resources and usable features, as the Service Blueprints stored in the Service Catalog and available testing and monitoring tools. The portal includes set of loosely coupled modules exchanging messages via a message/routing service bus, following microservice architecture. Using Apache Camel, this service bus supports communication either by message queues or via ‘publish/subscribe' model. Such microservices enable for the facility to handle (for example) authorization, auditing operations related to verticals and the service catalog, including catalog browsing, service ordering and service inventory related operations. To allow verticals to gain access to the framework, and hence to the facility, OpenSlice offers the web frontend, with two UIs. This web frontend is implemented in Angular and interacts with the OpenSlice backend API using an API proxy.

7.2.2 High Level Description of Protocol and Data Model

OpenSlice depicts the Network Slice as a Service (NSaaS) delivery model [30], in which each facility provides customized network slices to verticals on demand. Each vertical makes use of the given slice to meet their needs for trialing activities, setting up different use cases, and evaluating their KPIs under different network conditions. The high-level architecture that enables the NSaaS delivery model, in which the OpenSlice portal and service catalogue expose TM Forum Open APIs to verticals, allowing them to directly activate necessary operations for service ordering. The output of the Service Catalogue is derived from Facility Service Catalogue offerings hosted by the respective Service Orchestrators (SO). The service order is then forwarded to the SO on-site at the facility. The SO, which implements the 3GPP Network Slice Management Functionality (as both the Network Slice Management Function (NSMF) and the Network Slice Subnet Management Function (NSSMF), then instantiates the network slice through subsequent calls to the appropriate network function orchestrator (NFVO). The Network Orchestrator is coupled to an NFVO (NFV Orchestrator), which orchestrates VNF and network service instances with the assistance of a Virtualized Infrastructure Manager (VIM). The VIM manages the virtual resources on which those instances reside. The northbound interface is implemented by the NFVO using ETSI SOL 005. RESTful APIs (Application Programming Interfaces That Are Easy to Use)

7.3 Mechanisms Illustrated for Exposing NSaaS to Verticals

7.3.1 Supported TM Forum Open APIs in OpenSlice

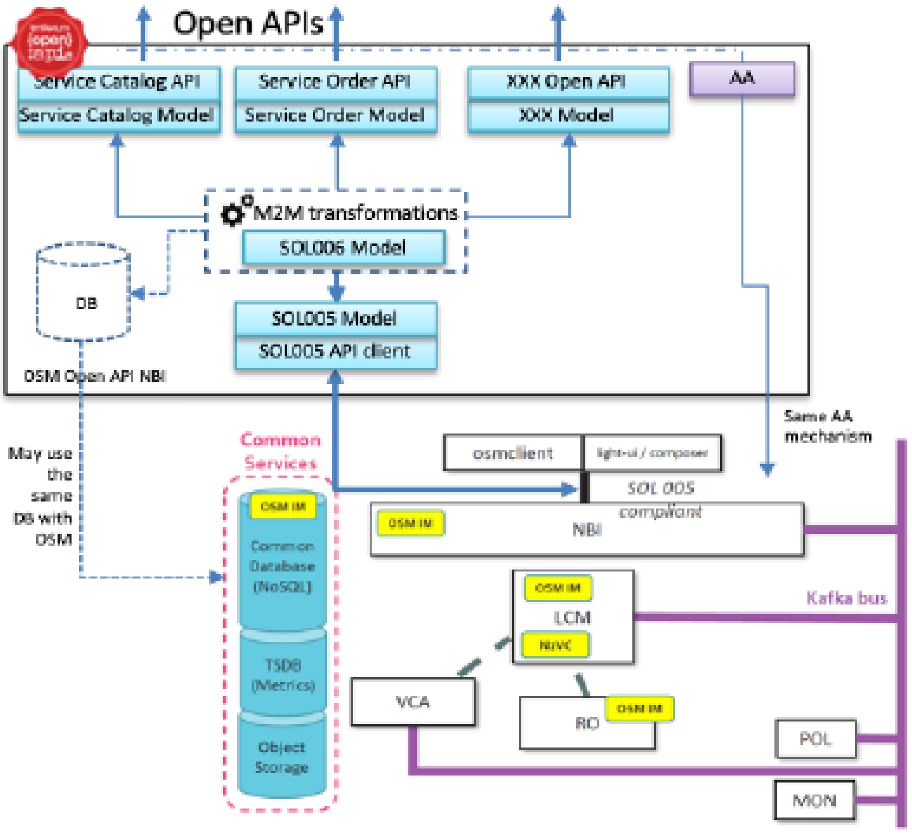


Figure 42: OpenSlice API Types of Exposure

7.3.1.1 Available APIs for Consume Services from OpenSlice

The Network Slice Provider offers TM Forum Open APIs, making them available for verticals, so they can invoke relevant NSaaS operations. To facilitate the handling of service catalogue, from SB design to vertical-triggered service ordering, OpenSlice makes use of the following TM Forum Open APIs [31]:

- Service Catalog API (TMF633), providing artifacts (e.g., models and dependencies) for the design of SBs, and capabilities for their lifecycle management (e.g., registration, deletion, updating, etc.) in the service catalog.

- Service Ordering API (TMF661), for issuing a service order. This order conveys the information required to instantiation parameters. In some cases, this instance can be modelled as a network slice subnet instance.

• Service Inventory API (TMF641), which defines standardized mechanisms for CRUD operations over the records providing run-time information about deployed slice (subnet) instances.

• Service Configuration and Activation API (TMF640), providing capabilities to allow the operation of a deployed slice (subnet) instance. This includes the ability to trigger lifecycle management actions (e.g., creation, modification, update, deletion) over that instance, and the ability to define rules to collect monitoring data from that instance (e.g., using threshold-based alarms or periodic notifications).

7.3.1.2 Available APIs for Integrating Services in OpenSlice

OpenSlice can consume services from 3rd parties via Open APIs. By Using the TMF632 Party Management Model is specifying the Organizations that can exchange items and other information such as:

- Import Service Specifications.
- Create a Service Order.
- Use the Service Inventory to query the status of the service ordered to the external partner organization.

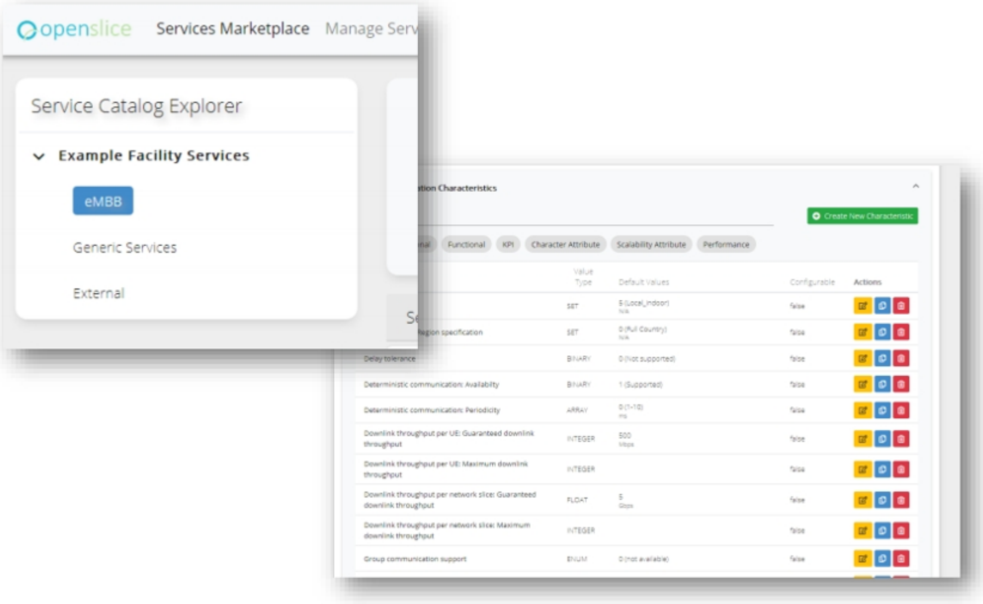
7.3.1.3 Summary of Supported TMF APIs

Supported TMF APIs in OpenSlice as of today are the following:

- TMF 620 - Product Catalog Management
- TMF 622 - Product Ordering Management
- TMF 633 - Service Catalog Management
- TMF 634 - Resource Catalog Management
- TMF 638 - Service Inventory Management
- TMF 640 - Service Activation and Configuration
- TMF 641 - Service Ordering Management
- TMF 666 - Account Management specification
- TMF 669 - Party Role Management
- TMF 629 - Customer Management
- TMF 632 - Party management

Supported APIs: http://portal.openslice.io/tmf-api/swagger-ui.html

7.3.2 Service Catalog and Service Blueprint Design



	Value Type	Default Values	Configurable	Actions
	SET	5 (Local_indoor) N/A	false	
Region specification	SET	0 (Full Country) N/A	false	
Delay tolerance	BINARY	0 (Not supported)	false	
Deterministic communication: Availabilty	BINARY	1 (Supported)	false	
Deterministic communication: Periodicity	ARRAY	0 (1-10) ms	false	
Downlink throughput per UE: Guaranteed downlink throughput	INTEGER	500 Mbps	false	
Downlink throughput per UE: Maximum downlink throughput	INTEGER		false	
Downlink throughput per network slice: Guaranteed downlink throughput	FLOAT	5 Gbps	false	
Downlink throughput per network slice: Maximum downlink throughput	INTEGER		false	
Group communication support	ENUM	0 (not available)	false	

Figure 43: Service Catalog/Specification Design

Referring to the Service offering two structures are defined to allow verticals to quickly bring their use cases into OpenSlice and subsequently into the CSP facility [32]. These are the:

1. the Service Blueprint (SB)

2. the Service Catalogue (SC).

The SBs announced by each facility site are registered and published into a single Service Catalog. This approach allows providing verticals with a unified marketplace, informing them about available service offerings in the entire facility. The Service Catalog offers various ways to access the defined Service Specifications, organized in categories (e.g., eMBB, Networking, edge, etc.). Verticals can browse the publicly available catalog and see offered services. When a vertical or a user cannot use a pre-existing SB for its test purposes, is able to create a new SB that can be based on a previous one. Details of each service are available, so that the vertical can understand underlying details, as well as the underlying attributes of the service offered. Also, any user defined attributes as well as ways to access the final operation service are also available. A service blueprint is describing a given network slice to be provisioned using NSaaS. This service template is a structured document that provides a complete description of a given network slice, including information on service topology and expected behavior. All this information is made from the Service Specification design. Any authorized vertical is in position to browse the SC, select a SB, fill it according to the service requirements, and issue a service order. Upon receiving this service order, the facility proceeds with the CFS-RFS translation, mapping the service order (CFS) into a network slice instance (RFS), deployed across one or more facility sites.

7.3.3 The Instantiated Network Service

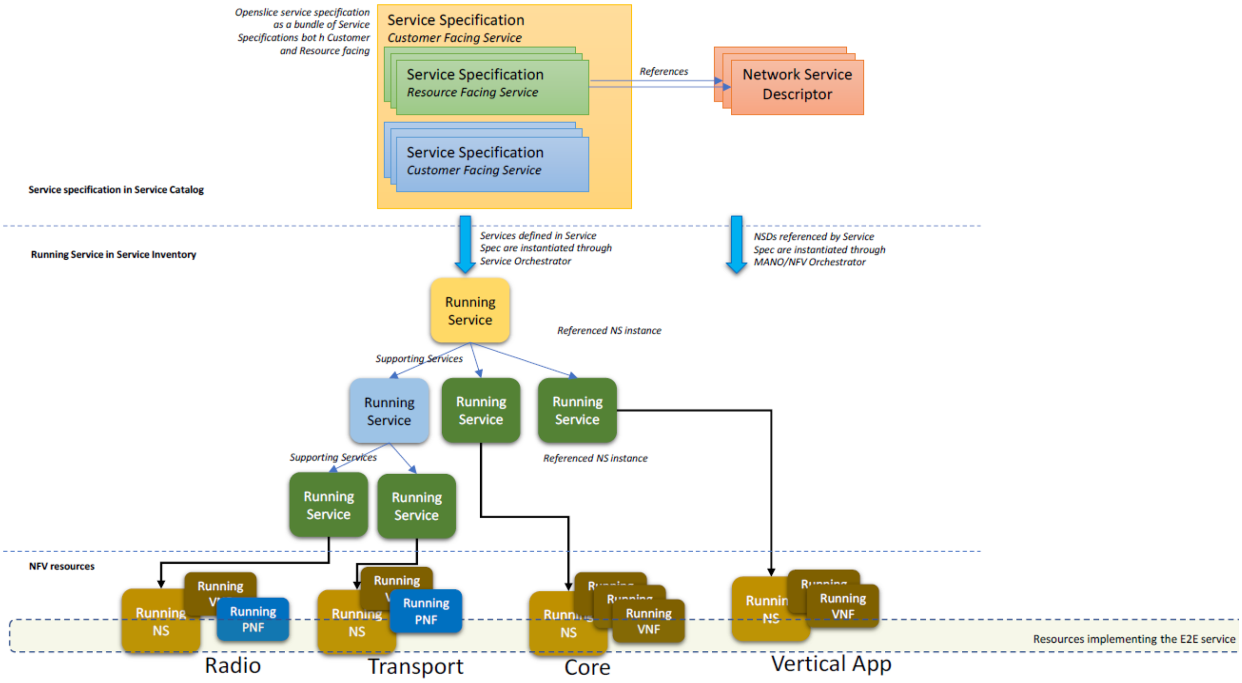


Figure 44: Service Specifications, Service Order and Running Services related with instantiated running Network Service.

The Verticals make Service Orders and OpenSlice instantiates the requested Service Specifications of the Service Order [28]. Running Services instantiated by OpenSlice, reside in OpenSlice Service Inventory. The picture displays how Service Specifications are related to Running Services and how Running Services relate with instantiated running Network Service. Figure depicts how Service Specifications, Service Order and Running Services related with instantiated running Network Service. A Service Specification can consist of several other services, called as a Service Bundle. It consists of RFSs having each RFS related with underlying NSDs. The Service specifications defined in the Service Order are instantiated through the Service Orchestrator. The NSDs referenced by each RFS Specification are instantiated through MANO/NFV Orchestrator. So, there is a top-level Service instantiated reflecting the user service, e.g., eMBB 5G slice. This for example may be decomposed and references 3 other Services which run and registered in OpenSlice Service Inventory. Other Running Services maybe further decomposed to other Services. This is done until each service is mapped to exactly one Running Network Service in the underlying Infrastructure. The Running NS is implemented by the VNFs and PNFs of the Communication Service Provider.

7.4 OpenSlice Deployment in Multi-Vendor 5G Facilities

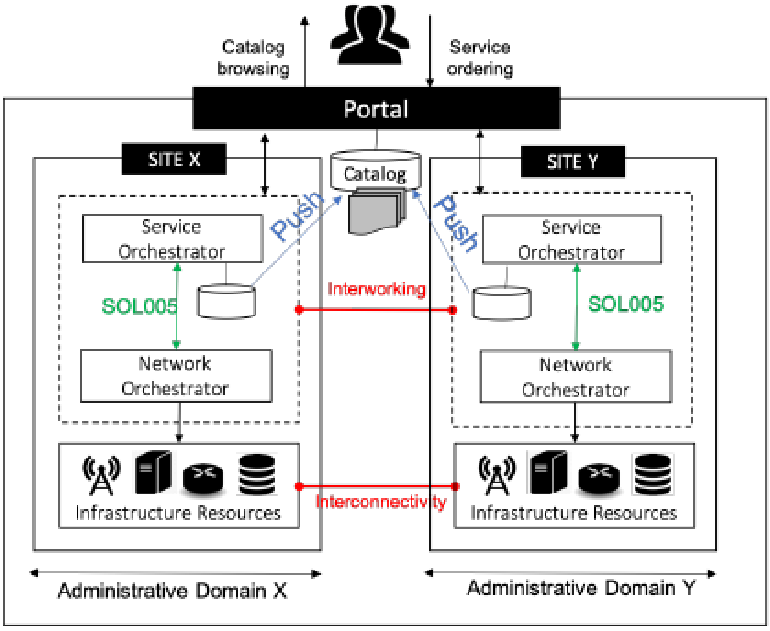


Figure 45: Position of OpenSlice Portal in a Multi-Site Deployment

To allow site-agnostic of network slices, and eventually the deployment of slices across two or more sites, Service Blueprint defines a common information model for the whole CSP facility [33]. Taking advantage of this feature, the facility collects the service offerings from the different sites, retrieving the SBs from their catalogs and publishing them all into a single, centralized service catalog. This new catalog allows the facility to provide the mentioned unified marketplace for verticals, facilitating their browsing and service orderings. Note that the responsibility of keeping the above catalog updated is up to every site. Every time the catalog of a given site suffers from any modification (e.g., on-boarding of a new SB, update of an existing SB), the corresponding network operator shall notify this change to the centralized catalog, using pushing mechanisms to that end. Every facility site operator designs their SBs, according to the features deployed on their administrative domains. For example, there are some sites that do not have edge computing nodes, and hence are unable to offer a uRLLC SB. The facility collects the SBs from the different sites and registers them into a single service catalog. Any vertical is in position to browse the service catalog, select the SB that best fits his needs, and issue a service order. This service order will be translated into a network slice instance, deployed across one or more facility sites. The OpenSlice provides a single entry-point for facility. Its mission is twofold. On the one hand, the portal support verticals customers taking the role of experimenters, in obtaining access to facility resources and available functionality (SBs, testing and monitoring tools) for the purposes of use case trialing and KPI validation. On the other hand, the portal allows exposing the facility to verticals as a unified platform rather than as an interconnection of individual administrative domains, hiding the mechanisms, protocols and technologies adopted in every site.

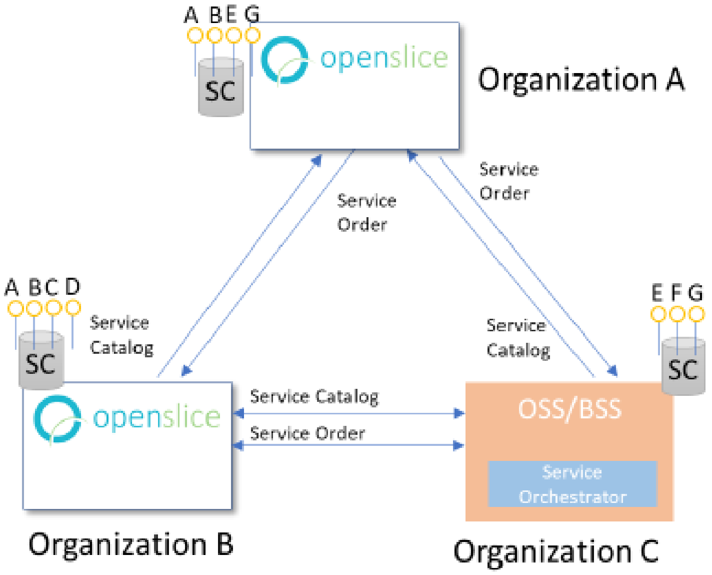


Figure 46: OpenSlice in a mesh connectivity for service exchange between Providers

A typical deployment across domains, involves today some typical components [28]:

- An OSS/BSS to allow customers access the service catalog and perform service orders.
- A Service Orchestrator (SO) component for executing the service order workflow.
- A Network Functions Virtualization Orchestrator (NFVO).
- The network resources configured by the NFVO/

TMF Open APIs are introduced not only for exposing catalogues and accepting service orders, but also implementing the East-West interfaces between the domains.

7.5 High Level Description from Service Order to End to End Service

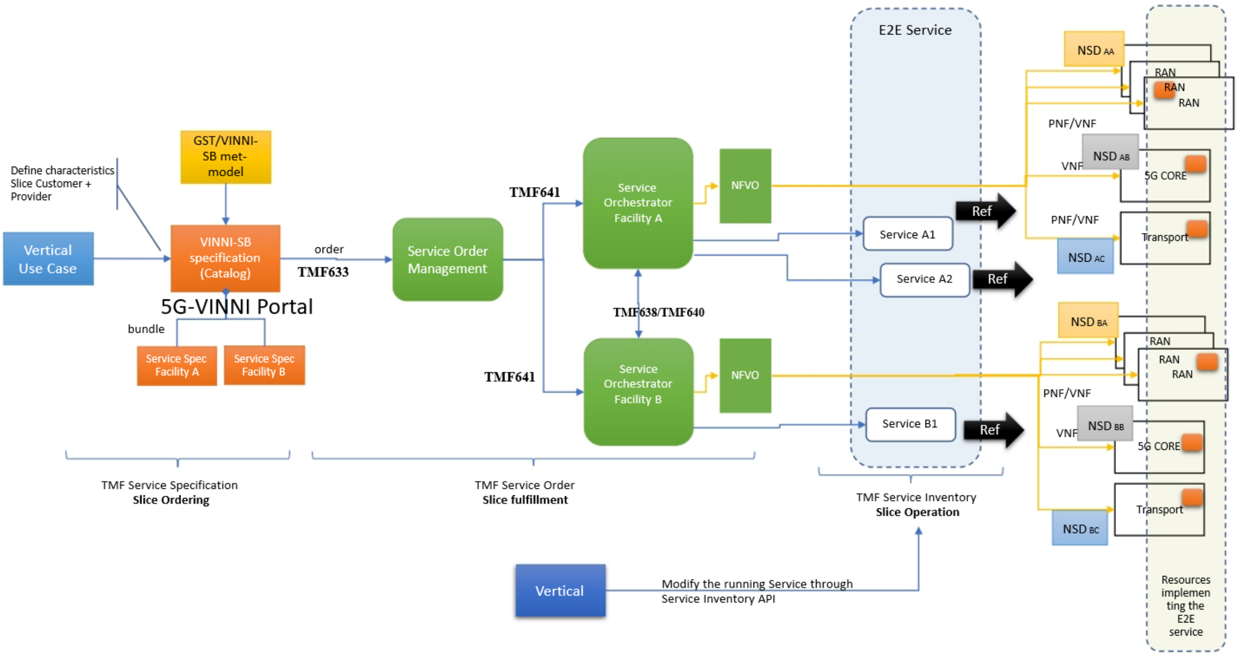
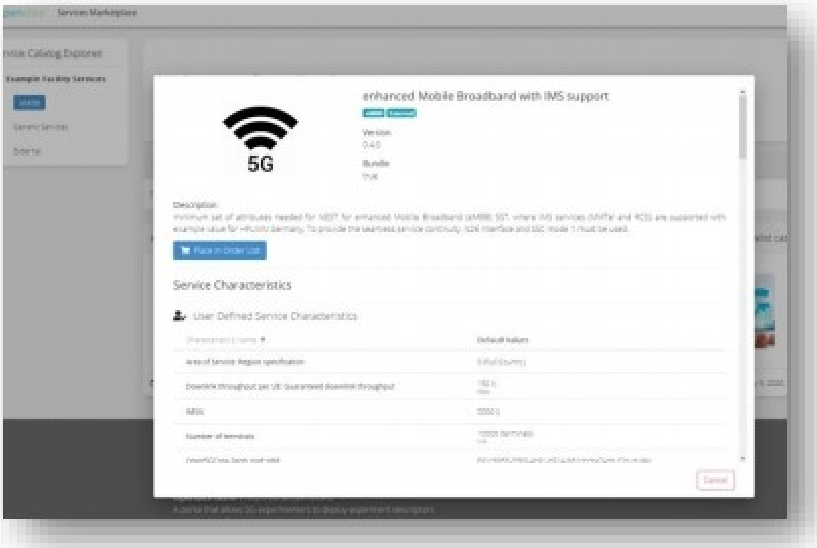


Figure 47: High Level Illustration from Service Order to E2E Service

The Procedure from Service order to End to End Service can be summarized to the following main steps [28].

7.5.1 Slice Ordering Phase



enhanced Mobile Broadband with IMS support

Version
0.4.0

Bundle
true

Description:
minimum set of attributes needed for NEST for enhanced Mobile Broadband (eMBB) 5G*, where IMS services (MMTel and RCS) are supported with example value for HPLMN Germany. To provide the seamless service continuity, S2X interface and SSC mode 1 must be used.

Service Characteristics

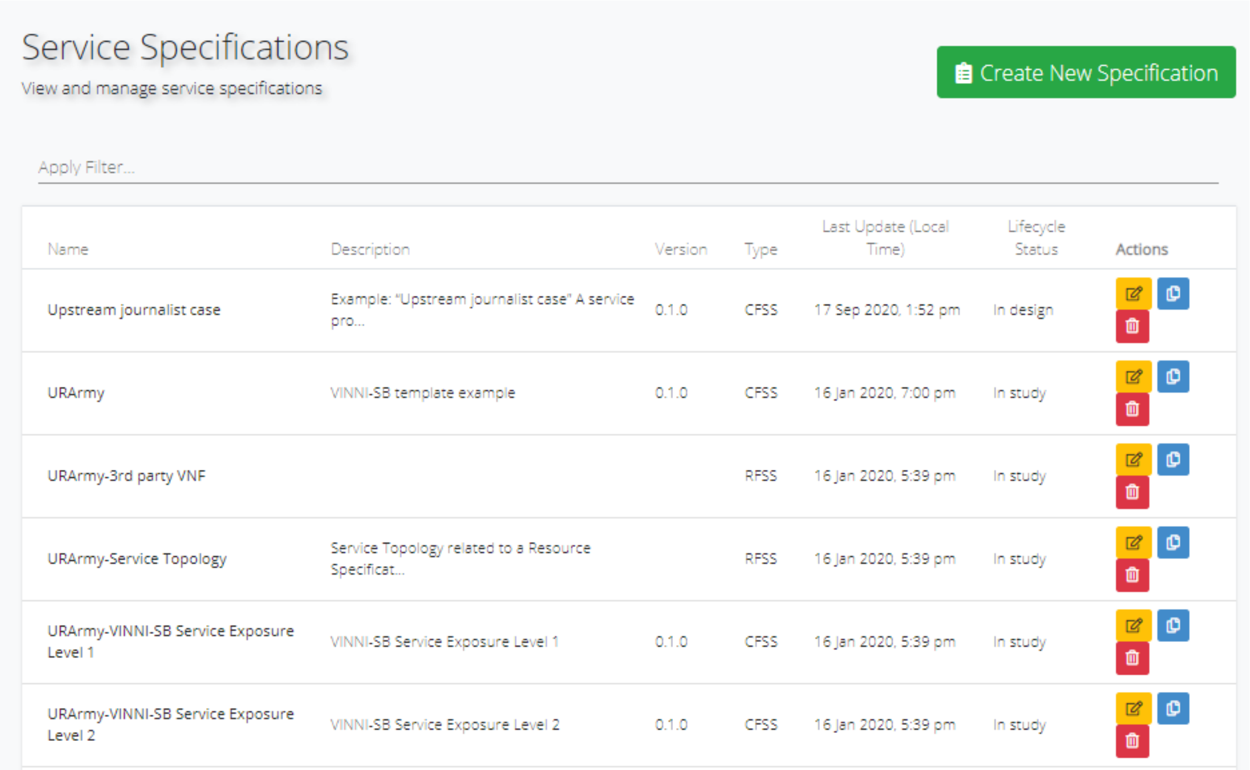
User Defined Service Characteristics

Characteristic name	Default Values
Area of Service: Region specification	0 (Full Country)
Downlink throughput per UE: Guaranteed downlink throughput	150 ()
MBSI	2000 ()
Number of terminals	10000 (terminals)
[illegible]	[illegible]

Figure 48: Slice Ordering

The Slice Ordering is initiated from a 5G Vertical (a NS Customer), by ordering specific Service Specifications that are related with Network Services. These Service Specifications are located in a Service Catalog of the provider. Both Specification and Service Catalog follow the model of (**TMF633**), which provides artifacts (e.g., models and dependencies) for the specification of Services. We have two types of Service Specifications: Customer Facing Services (CFS) and Resource Facing Services (RFS). CFSs are exposed to the catalog and users can order them. RFSs are internal and related with the underlying resources. Assuming these service specifications follow the GST model, the derived NESTs are modeled as TMF Service Specifications. Actually, from a GST meta-model a TMF model is created. A NEST (a Service Specification) usually is a Service Bundle. It consists of RFSs having each RFS related with underlying NSDs (VNFs) for radio, core, or other applications. For example, a service spec for a 5G Slice on Country Greece and Location Athens, might consists of RFSs that will configure via VNFs and PNFs, RANs in Athens and create a new 5GCore while configuring an existing transport network. Thus, the Customer requests the Service Spec and creates a Service Order. Service Orders are modeled with the **Service Ordering API (TMF641),** which allows issuing a service order, and includes selected Service Specifications as well as instantiation parameters. (e.g., Network Slice region, QoS, Number of UEs, etc.)

7.5.2 Slice Fulfillment Phase

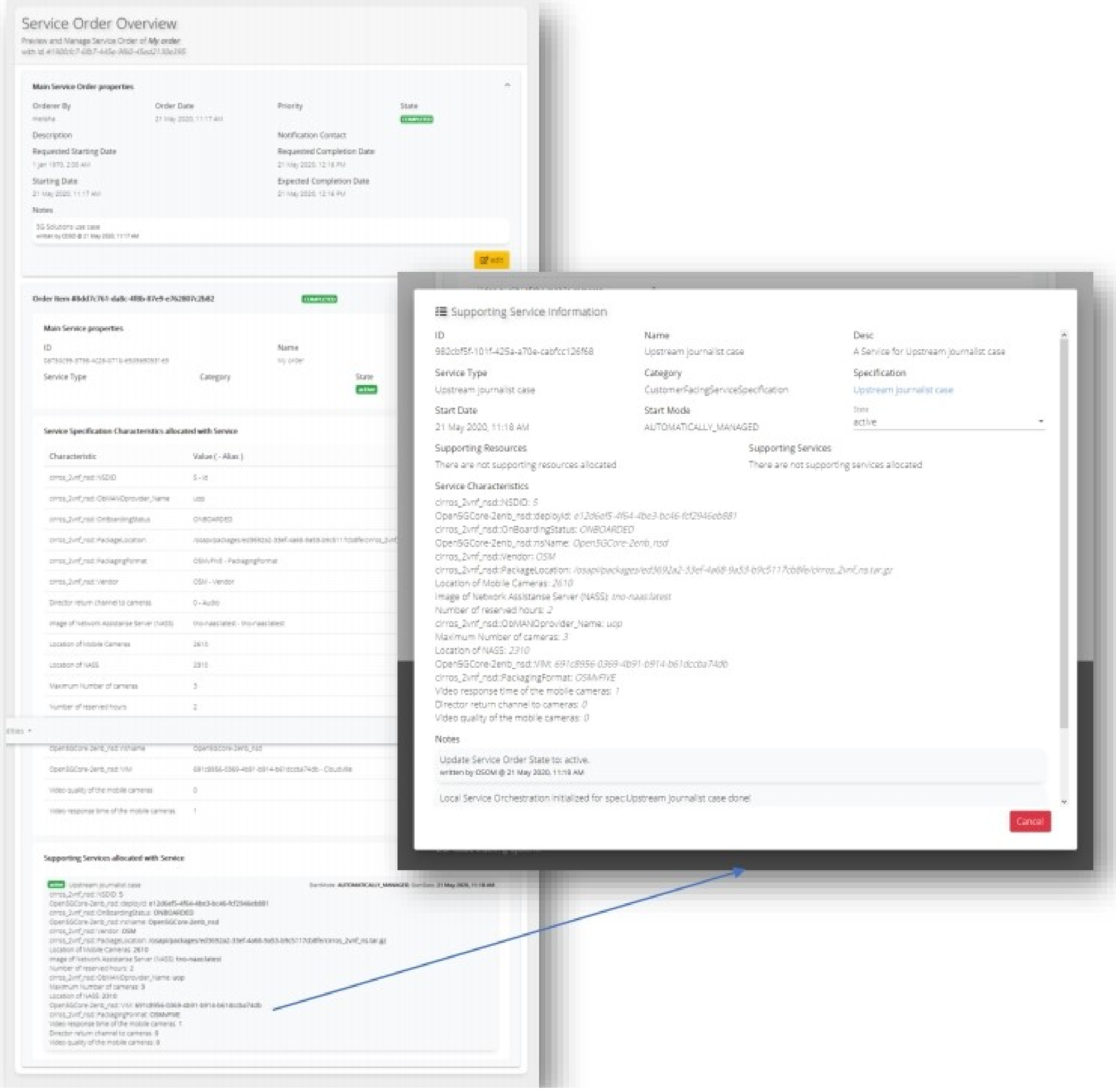


Name	Description	Version	Type	Last Update (Local Time)	Lifecycle Status	Actions
Upstream journalist case	Example: "Upstream journalist case" A service pro...	0.1.0	CFSS	17 Sep 2020, 1:52 pm	In design	
URArmy	VINNI-SB template example	0.1.0	CFSS	16 Jan 2020, 7:00 pm	In study	
URArmy-3rd party VNF			RFSS	16 Jan 2020, 5:39 pm	In study	
URArmy-Service Topology	Service Topology related to a Resource Specificat...		RFSS	16 Jan 2020, 5:39 pm	In study	
URArmy-VINNI-SB Service Exposure Level 1	VINNI-SB Service Exposure Level 1	0.1.0	CFSS	16 Jan 2020, 5:39 pm	In study	
URArmy-VINNI-SB Service Exposure Level 2	VINNI-SB Service Exposure Level 2	0.1.0	CFSS	16 Jan 2020, 5:39 pm	In study	

Figure 49: List of Service Specifications, both RFSs and CFs

The slice fulfillment phase begins when the customer triggers a service order from the service catalog. In a service order, the customer provides a completed specification of the slice instance he wants, including information on slice topology (possibly extended with 3rd party VNFs) and slice attributes (filled in with values fitting use case requirements). To do that, the vertical makes use of the Service Ordering API. The issued service order is then captured by the Service Order Management (SOM), which propagates it towards the corresponding Service Orchestrator(s) again via TMF641 API calls. Next, the service order is processed. This processing consists in translating the received service order (CFS, handled by the Service Orchestrator) into a set of resource requirements for the network slice to be instantiated (RFS, handled by the NFVO). To successfully achieve this translation, Service Orchestrator and NFVO exchange information relying on ETSI SOL005 capabilities. Once this translation is completed, the NFVO allocates the slice instance, instructing the VIM for that end. At this stage we have the E2E network service running involving Radio, 5G Core, Transport or any other desired Network Service (e.g., a firewall).

7.5.3 Slice Operation Phase



Service Order Overview

Preview and Manage Service Order of My order

Main Service Order properties

Orderer By: manisha | Order Date: 21 May 2020, 11:17 AM | Priority | State: COMPLETED

Description | Notification Contact

Requested Starting Date: 1 Jan 1970, 2:00 AM | Requested Completion Date: 21 May 2020, 12:16 PM

Starting Date: 21 May 2020, 11:17 AM | Expected Completion Date: 21 May 2020, 12:16 PM

Notes: 5G Solutions use case

Main Service properties

Service Specification Characteristics allocated with Service

Supporting Services allocated with Service

Supporting Service Information

ID: 982cbf5f-101f-425a-a70e-cabfcc126f68 | Name: Upstream journalist case | Desc: A Service for Upstream journalist case

Service Type: Upstream journalist case | Category: CustomerFacingServiceSpecification | Specification: Upstream journalist case

Start Date: 21 May 2020, 11:18 AM | Start Mode: AUTOMATICALLY_MANAGED | State: active

Supporting Resources: There are not supporting resources allocated

Supporting Services: There are not supporting services allocated

Service Characteristics

cirros_2vnf_nsd::NSDID: 5
Open5GCore-2enb_nsd::deployId: e12d6ef5-4f64-4be3-bc46-fcf2946eb881
cirros_2vnf_nsd::OnBoardingStatus: ONBOARDED
Open5GCore-2enb_nsd::nsName: Open5GCore-2enb_nsd
cirros_2vnf_nsd::Vendor: OSM
cirros_2vnf_nsd::PackageLocation: /osapi/packages/ed3692a2-33ef-4a68-9a53-b9c5117cb8fe/cirros_2vnf_ns.tar.gz
Location of Mobile Cameras: 2610
Image of Network Assistanse Server (NASS): tno-naas:latest
Number of reserved hours: 2
cirros_2vnf_nsd::OSMANOprovider_Name: uop
Maximum Number of cameras: 3
Location of NASS: 2310
Open5GCore-2enb_nsd::VIM: 691c8956-0369-4b91-b914-b61dccba74db
cirros_2vnf_nsd::PackagingFormat: OSMvFIVE
Video response time of the mobile cameras: 1
Director return channel to cameras: 0
Video quality of the mobile cameras: 0

Notes

Update Service Order State to: active.
written by OSOM @ 21 May 2020, 11:19 AM

Local Service Orchestration initialized for spec:Upstream journalist case done!

Figure 50: Service Order and Service Order Management

As described, TM Forum Open APIs may allow customers not only to interact with the service catalog but also to consume the exposed capabilities. In the slice operation phase, the slice is already instantiated, and can be made available for operation. During this phase, the customer keeps track of the status of the slice instance, making use of the Service Inventory API (TMF638), which defines standardized mechanisms for CRUD operations over the records providing run-time information about deployed slice instances.

7.6 OpenSlice Support for Testing as a Service

The Service Blueprint has the option to include Testing as a Service (TaaS) [30]. For example, by including relevant test scripts for TaaS. The test scripts will represent specific Test Cases (TCs) that are targeted at stressing specific aspects or KPIs of the Service. It is important to understand that there are two types of TCs: 1) TCs for validating fundamental network KPIs like throughput, delay, and 2) TCs for validating the service of the vertical customer that might include the vertical's application. While the former can be offered as part of the service design with minimal configuration, the latter might require significant work. Such services can be consumed both as a human-driven or an automated interaction. TaaS consists of a set of testing tools that can be deployed, configured, and automated through a set of offered web services. A typical example of testing tools is traffic generators, that can emulate realistic traffic and protocols. The offered services allow the vertical customer to:

- Onboard specific drivers for automating vertical applications and use cases.
- Create and execute individual test scripts for automating the tests or the experimentation.
- Create and execute test campaigns, i.e., batch of test scripts that can be executed on multiple target infrastructures.
- visualize logs and results through the offered visualization systems.
- allow the vertical to develop customized visualizations.

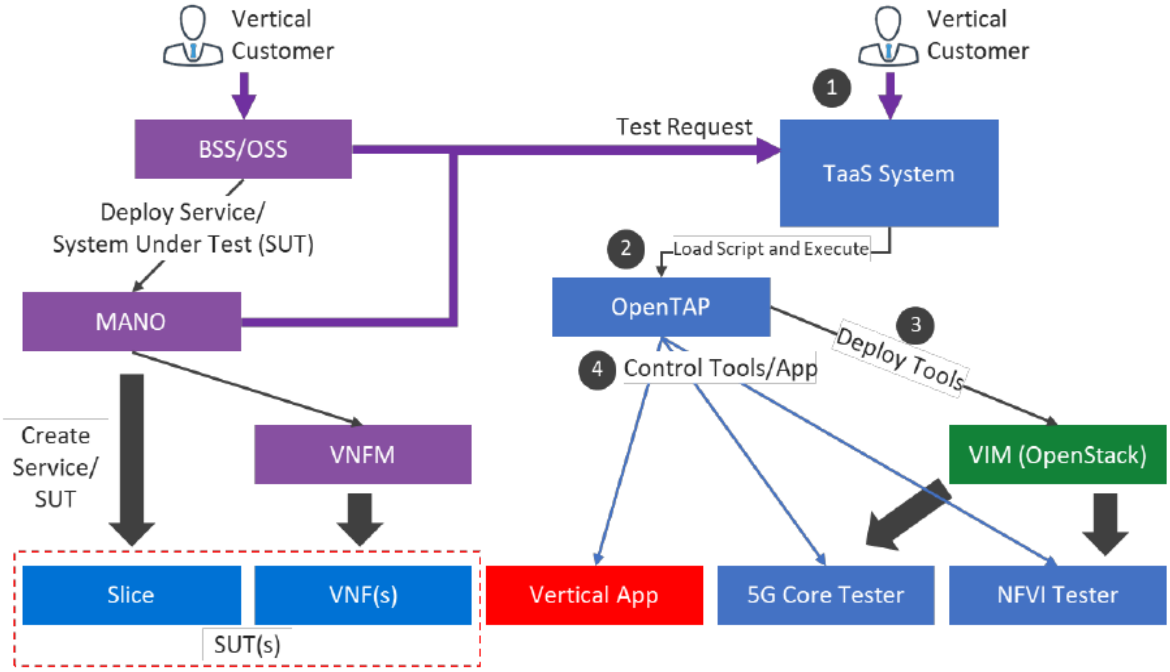


Figure 51: Consuming a TaaS Service

In Figure, two examples are depicted for how to consume the testing service. The first case is for service validation. While deploying the Network Service, a set of test cases can be automatically requested by the BSS/OSS and executed after the deployment in order to verify the health or performance of the service. In this case the test cases are programmatically requested via API calls (TaaS APIs are not yet published) to the TaaS. The other way of consuming the service is using web services (TaaS web services are available and will be supported for 5G-VINNI Release 2), that allow manual execution of the tests. The manual execution allows for a more comfortable way of configuring the tests to perform more exploratory experiments. After the tests is requested (1), the test scripts present in the TaaS repository are loaded and executed on OpenTAP, an open-source technology that is at the heart of the TaaS system. OpenTAP allows to programmatically deploy tools e.g., in an OpenStack cloud, and configure them to target the newly deployed service.

8. FUTURE CHALLENGES AND RESEARCH DIRECTIONS

8.1 Different Understanding and Emphasis of Each Organizations

Table 1: Summary on the Similarities and the Differences Between SDOs' Visions

SDOs Modelling Aspects	Provision Models			NS Service Request	Network Domain				Main Contribution
	SaaS-like	PaaS-like	IaaS-like	Service/Slice profile	RAN	Core	Transport	General	
ETSI		✓						✓	Generalized NS architecture and its associated workflows
3GPP	✓			✓	✓	✓			Information/data model for NS provisioning and management on RAN and Core network
GSMA	✓			✓	✓	✓	✓		GST/NEST
IETF		✓					✓		Technology-Independent information model for NS management via a network-agnostic interface

Based on the above diverse definitions of network slicing, it is easy to find that the different organizations have a different interpretation of network slicing principle and have their focus on each one particular aspect. In 3GPP, network slice is implemented in RAN and CN for a mobile network. Network slice instance can be interpreted as a collaboration of CN and RAN and is something of a customized function in the core network. However, no standardization has been provided from the transport layer between the RAN and the CN. The IETF wants to describe a more comprehensive range of implemented end-to-end network slicing for both mobile networks, but also other networks modes which do not include both RAN and CN. Therefore, the transport layer is considered as a part of network slice instance and thus can be sliced as well. This may request an orchestration of several domain technologies in the transport networks and a unified information model to expose their related technical capabilities to verticals. While ETSI focuses greatly on how the NFV architecture can be applied efficiently to the network slicing rational, GSMA describes network slicing in a more generalized and business manner. The various definitions given by different organizations are intended to provide a detailed and overall standardization for network slicing addressing all its different aspects. According to the survey [34], the NS provision model through the transport layer is unclear at the IETF. Until now, three models are being developed for the NS Provision: The Software as a Service (SaaS), the Platform as a Service (PaaS), and the Infrastructure as a Service (IaaS) [35]. GSMA adopted the SaaS model, which was pioneered by 3GPP for both the request and provision of NS services. In the PaaS model, the tenant asks for the full set of Nodes and interconnection and connectivity configuration data. This is consistent with the goals of the ETSI Next Generation Protocol vision. Unlike cloud-based PaaS and SaaS models, IaaS tenants have control over the underlying infrastructure and can specify their resources and the underlying configuration. Since this approach does not align with the network-wide best practices for network slicing, it is not recommended it. However, can be considered as an interesting initiative for further research as no effort was made to validate the applicability of the model. Table 1 summarizes the various SDOs' visions.

8.2 The IP Perspective of Network Slicing for “Network as a Service”

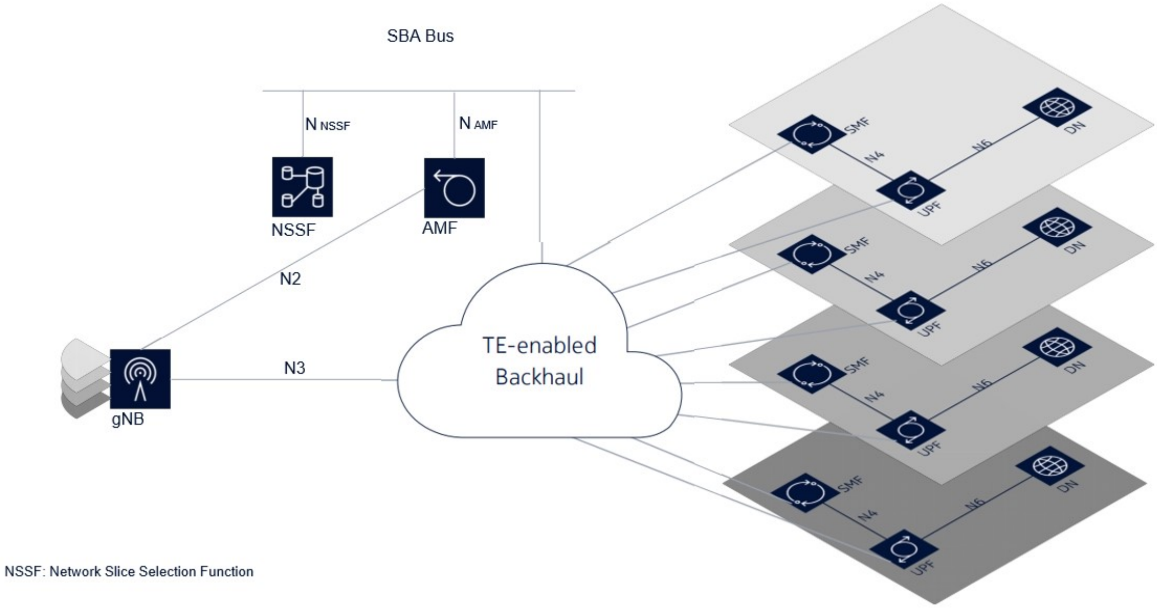


Figure 52: Service Aware Backhaul Network (Source: Nokia)

The concept of slicing a network into virtual containers or virtual elements is not new for IP networks [36]. Similar implementations already exist like VLANs. The difference in network slicing in 5G is basically that a slice is being instantiated all the way from the cell phone to the core elements, so this means that the cell phone is to be provisioned for using that slice, the radio needs to be provisioned for using that slice, and the transport layer needs to be provisioned for using that slice. Further to the transport layer there is not a standardization on how the slice should be instantiated in the transport layer. So, the whole concept of slices is to create a specific infrastructure to serve a set of UE, demands that might need to have some specific functions dedicated for that slice, in order to provide some specific isolates that are different slices. There is a slice that demands some lower latency or a slice that demands some higher throughput or a combination of both in such a way from an IP perspective the infrastructure ends up with a classical traffic engineering problem. There are some traffic flows demanding low latencies, some traffic flows that demand higher throughput. All these variables can be concerned in terms of providing isolation in an IP transport network and that is nothing different than a traffic engineering problem. So, by choosing the right traffic engineering protocol several traffic engineering mechanisms can be applied like segment routing or MPLS TE in a more centralized manner. One thing that might be a little different is probably not something exclusive to 5G but it's something that 5G will make bigger or scale to much larger, is the fact that all of these elements like SMF UPF AMF NSSF they are not instantiated by the use of physical hardware they are virtual elements. So, they are going to be installed as either a VM or a container-based application on the data center that needs to go through fabric needs to go through a data center interconnect. So eventually what is needed is to extend that traffic engineering domain inside that data center environment. There are several ways of solving that scaling problem, but the biggest challenge is coming upon start building more distributed infrastructure in terms of the core to get closer to the user. That number of mini data centers will be much larger and then the cross-domain orchestration not only across the data centers but also at the radio and at the other core elements in a 5G network is going to be a key to realize this vision of network slice or network-as-a-service over 5G. Eventually, various networks have been built for a mobile operator or for a radio where the radio is the customer. So, a dedicated infrastructure is built to serve this customer which is the radio network that needs interconnection to the core.

8.3 ORAN Architecture for Network Slicing

8.3.1 ORAN Protocol Stack Disaggregation

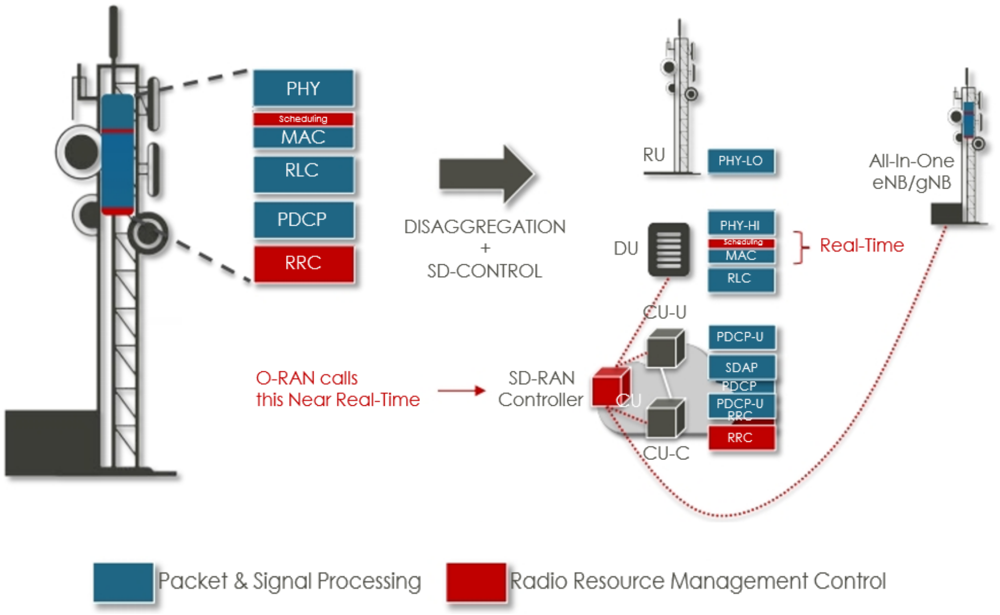


Figure 53: RAN Disaggregation for Network Slicing

The open radio access network (ORAN) [37] protocol stack has two distinct components where radio Resource Management (RRM) control is conducted. The RRC Layer and the Scheduling which is a component of MAC Layer. The disaggregation of RAN [38] protocol stack leads to an infrastructure composed of the Radio Unit (RU) for part of the physical layer, the Distributed Unit (DU) located closely to RU running RLC, MAC and parts of the PHY Layer and the Central Unit (CU) running RRC and PDPC layers and capable of controlling the operation of many DUs. An SDN based control mechanism can be applied separating CU into CU-U and CU-C orchestrated by an SD-RAN Controller with standardized interfaces between the controller and components. In the future this disaggregation can be avoided by introducing a potentially compliant all-in-one base station. O-RAN calls this as the Near Real-Time RAN Intelligent Controller (Near-RT RIC). The reason called near real-time is because in the wireless Channel network states are changing rapidly. There is only a certain number of Radio Resource Management operations thus it is not needed to follow all those changes on as rapidly as they occur. Just a filtered version is sufficient. The Near Real-Time RAN Intelligent Controller is logically centralized and it's at the edge. In the same rational the DU component can be referred as real-time.

8.3.2 RAN Controller

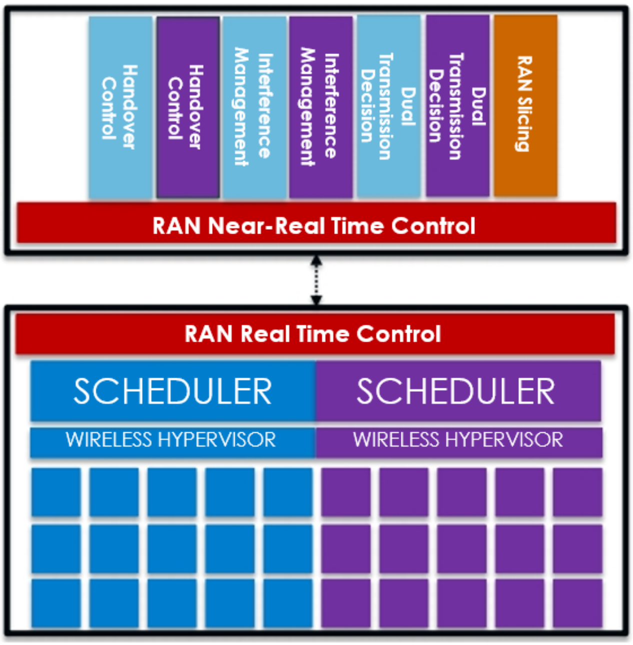


Figure 54: RAN Controller

The RAN controller consists of the near-real-time controller that has a bunch of RRM control applications and the real-time controller. The real-time Controller is a central scheduler having probably several schedulers on it and as far as the network state is concerned, these are the real time controls requiring instantaneous QCI values. The near-real-time controller consist of control mechanism would probably need time average QCI Values. The scheduler is responsible for several procedures taking place. First decides on which users to serve at a given time. Also decides on how much of the spectrum to give to each of the scheduled user. Finally, it does the actual mapping between the resources needed to be allocated with the number of resources allocated to a particular user. So, slicing of the RAN comprises the disaggregation of the scheduling operation and the virtualization of the available physical resources in such a way that the scheduler maps the physical to the virtual resources. The resulting architecture will compose of dedicated schedulers with access only to their own virtual resources, serving a number of particular users based the type of service they are requesting. At the end, collectively a mapping of a virtual to physical resource block is performed on a transmission time interval (TTI) level. Then, the configuration of these slices is provided from the near real-time controller through a slicing application installed on the (Open Ran Intelligent Controller) RIC. The configuration is provided almost in real time and this architecture composing a slice-oriented network where specific RRM mechanisms like specific handover and interreference management algorithms can be used per slice. Overall, can be considered as many virtual base stations dedicated to different slices in a composite physical hardware.

8.4 Extending Network Slicing at the Edge

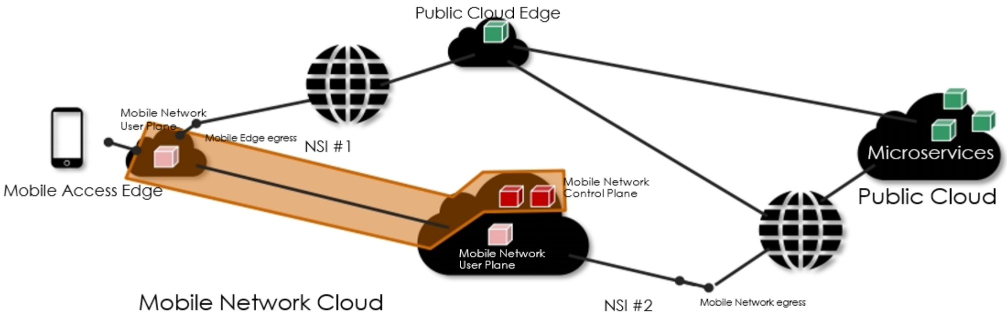


Figure 55: UPF at the edge

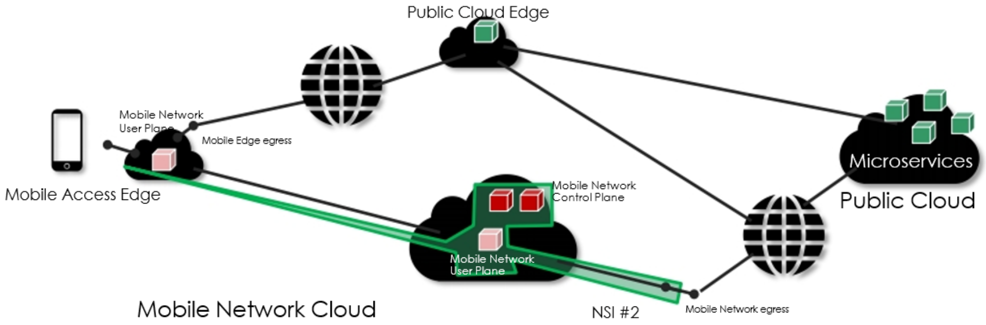


Figure 56: UPF in the Central Mobile Network Cloud

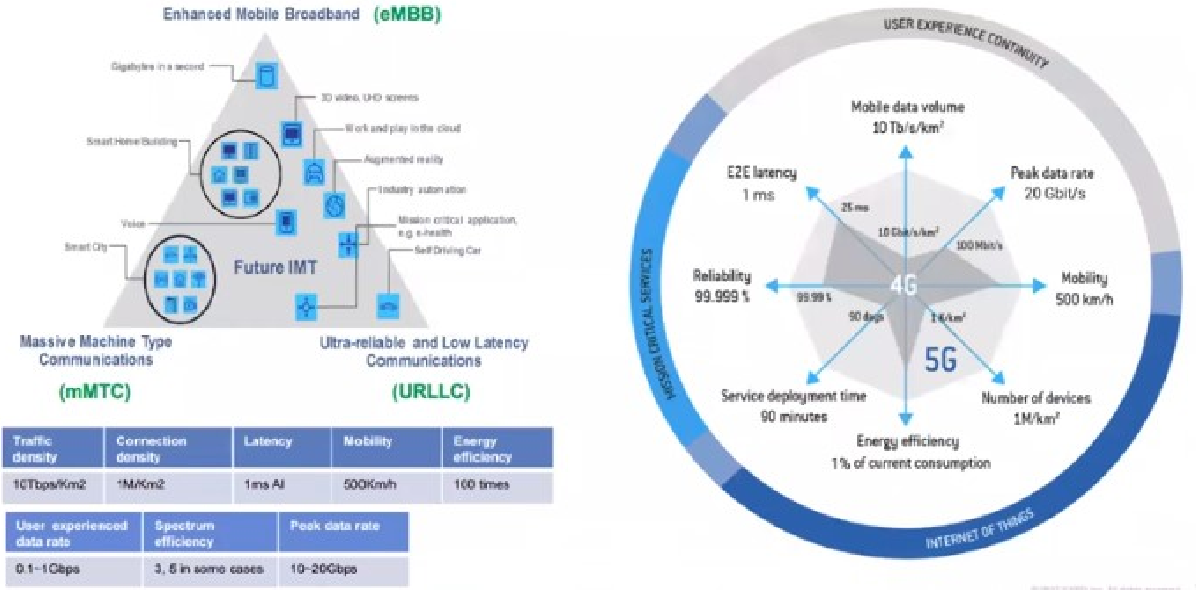
Referring to the 5G Edge, the mobile network can be defined as the distributed Network cloud. So, edge in high level is where public Cloud meets the network Cloud. In a scenario where the UE tries to reach an Over-the-top (OTT) application in the public cloud, should at first establish a connection or path through the mobile network (which includes the RAN) and the core network and through a mobile network egress to access the internet and finally to reach the public cloud where the OTT is hosted [39]. By disaggregating the OTT applications into microservices, the public cloud can move those microservices closer to the end user and afterwards to be pushed to the Enterprise locations. In parallel, operators are going through a transformation of their network into network clouds. The optimal solution as far as the end-users concerned is to host a microservice in the Mobile Network cloud. This solution might be not feasible as the public cloud and the Telecom Operators cannot agree on co-hosting. By introducing the network slicing rational, the operators will programmatically create network slices. There could be one slice where the user plane function is located in a mobile network cloud closest to a public network cloud where a microservice is requesting for a specific service and another network slice where the user plane and the control plane is in the central Mobile Network Cloud [40][41]. The benefit of this architecture which is feasible via the NFV and SDN is that the OTT application is becoming a tenant to the Telecom Operator's cloud and as a tenant asks for a particular service type through a particular network slice. So effectively the two clouds need to establish a communication where the public Cloud asking for content based on the type of service. This communication would be optimal if there was an end-to-end management of the whole operation.

8.5 Need for Slice Isolation

Slice isolation [42] is typically an important requirement in the case of multiple slices sharing the same infrastructure. That isolation means that the performance of each slice must not have any impact on the others. There are two aspects for the need to enhance the network slice isolation. The first is slice security. In case of a cyberattack or a fault occurrence, only the target slice will be affected. The second is slice privacy. The private information related to each slice its state and its traffic are not shared among the slices. The next step is to define the rules and the mechanisms to build and manage network slices by applying the principle of network function virtualization and the principle of software-defined networking.

8.6 Operation and Service Assurance Perspective for Success of Network Slicing

8.6.1 Emerging Needs for the Network and the Operations



Traffic density	Connection density	Latency	Mobility	Energy efficiency
10Tbps/Km2	1M/Km2	1ms AI	500Km/h	100 times

User experienced data rate	Spectrum efficiency	Peak data rate
0.1~1Gbps	3, 5 in some cases	10~20Gbps

Figure 57: 5G Functional and Operational Requirements

The demanding needs of the architecture of 5G overstretches the network and the operations and rely heavily on new technologies like NFV and SDN and cloud-based architecture [24]. The user Experience by adopting to the 5G must deal with two more fronts. The IOT which is not about content to the user, it is more of those users and the things that the users contacting the Network for services and for further components that are required to make the whole service works. Then there is another set of services that basically require real-time interaction between components in the network and devices like connected cars. Also, there is a demanding need for 5G to meet the requirements for additional emerging Services which also overstretches the network. Comparing with 4G and with the architectures before, the user and the whole network was optimized to provide a predefined set of services. In the same rational the aim of the service provider during the span of control was to provide that set of services. In 5G the IOT connected car and these Emergency Services are becoming more complicated because not only the service provider, but other stakeholders have access or influence on the service as well. In 5G this set of services at a high level are categorized into three main groups or three major use cases. The enhanced mobile broadband which requires high speeds in gigabytes for providing services like 3D Videos. The reliable low-latency Communications (URLLC) with use cases requiring low latency and real-time aspects such as self-driving cars, industrial automation and remote surgeries. The last one is massive machine type communication which is basically focusing on IOT like devices at home and monitoring aspects for services related to Smart Cities. Intuitively with all these beneficial features also expand the thread surface against the network.

8.6.2 Emerging need for the Operators to deploy 5G Networks.

The potential to deploy several network slices to serve diverse facilities is an emerging requirement for service providers deploying 5G networks. The network operator will then determine whether it needs a single slice or several network slices. For instance, within a single slice, an operator may support service types A and B, or it may provide service type A in slice 1 and service type B in slice 2. The UE must register in order to obtain access to a service run by the network. It could be beneficial to have a registration process that can accommodate a multitude of different deployment options to satisfy the different slice and service offerings in order to retain flexibility for the network operator. Therefore, a mechanism and method of registering a UE with a network service or a network slice or more than one is required to address the shortcomings of 4G, where there was effectively only one network slice with which the UEs were registered.

8.6.3 Aspects in Management and Operation Procedures

This whole set of knew services requires different mechanisms in terms of management and operation procedures and different treatment from the Customer perspective [26]. The are some main aspects in the architecture of 5G required to handle the emerging needs of network slicing. An aspect will be to reduce the life cycle of creating these services. The provided services need to be created quickly and should be fully operational for the required service. Another aspect is focusing on SDN. A software-defined network can control how the network behave and optimize itself for a specific service. It is not one-size-fits-all IP network as it is flexible to define the protocol and the network behavior by optimizing it for the service. Also, an aspect is the network virtualization which gives the capability not to deploy a physical function with the long lifecycle of deploying this function but pieces of software that can be chained together to provide the service. The architecture is starting to broaden itself as based on the specific service, a customer is able to choose and deploy the right software, the right function from the right vendor and with the right characteristics to actually fit this specific service. Another aspect that coming to mobility specifically is the cloud RAN and the new architecture for the central office. The traditional central office was mainly lots of equipment from different vendors with different operation, different expertise, and different skills. Cloud based RAN requires one type of skill where every customer needs only one type of skill to maintain it. This maintenance can be also taking place remotely as it has the flexibility to add functionality to it without adding complexity in maintaining this cloud RAN. The aspect of Mobile Edge Computing is going to enable a lot of services that require massive processing like AR like VR augmented reality. That will require the ability to provide processing power as close as possible to the user. Referring to the wireless part where the massive MIMO and beamforming make sure that the latency and the capacity is the required one for the required services.

8.6.4 Network Slicing Architecture Considerations

So, a little deeper into the network slice is basically an inter end network and not just a network but also the function associated with this network that supports a specific service and is optimized to this specific service based on the SLA with the customer. Has its own behavior and security aspects. Also, there are certain policies are associated with this network when the network has a problem or if the network needs to scale. The so called as the horizontal aspect of the slice is more than partitioning a piece of a physical Network where all the entities are managed the same way and optimized for the same Services. The network is separated to actual slices and is not being optimized just for the service but also operationally for that service. This was not possible until NFV and SDN because SDN gives the flexibility to build protocols and behavior for the network that is specific to the slice and also NFV is enabling to provide virtual network function and service chain that is specific to this slice. It's an infrastructure where a whole different network service chain is being built and optimized for this service.

8.6.5 Infrastructure Testing Considerations

Once the Network Slices have been created there is another aspect in terms of testing and monitoring. Testing the cloud and testing the network slice is extremely important. The network used to be built in the test lab is like a component in a bigger network where the equipment type and the traffic from the simulator are basically the same. So whatever behavior or performance metrics provided from the lab should be the same as scaling it with real traffic. The expected result in such case should be almost the same but a network slice is created virtually in the cloud and the components can be from different vendors and the requirements can be different, and the scalability also can be different. So, the Quality Assurance department needs to test the infrastructure in the cloud in the exact behavior and the exact situation that it would be used.

8.6.6 Complexity Demands Scaling in the Network Maintenance

The scaling of the complexity as the number of devices increases leads to the increase of the maintenance procedures along with the number of slices. Resulting to the previous event chain the Edge data centers need to be increased as well. This leads to a demanding need of an orchestration in the network. The devices are communicating with the entire network and some of them can be split when a slice is instantiated. So new instances spun up with day zero configurations. With virtualization is possible to provide embedded configuration in these instances which when they come up, the hardening process setup be able to initiate as well as the integration with the FCAP solution. They should be able to be logging based on the customer standards and they should have good vulnerability protection.

8.6.7 Network Maintenance Aspects

The existence of many slices instantiated in End-to-End isolated environments has to be done carefully as it may harden the network infrastructure. The validity of the configuration in the routers forwarding the traffic and depending on the slices should be executed continuously. So, hardening process setup needs to be defined in the packet core where the configurations are constantly being checked for gaps or for a proper identity and access management. The identity and access management are two important factors as a lot of vendors will access one Operator's network spanning to RAN vendors from MID-Hall vendors. This access needs to be managed in order to be able to revoke that access if needed. FCAPS fault configuration accountability performance security in production ensures that if a device is going out of capacity to notify before happens and take actions accordingly in order to avoid disrupt the Quality of Experience. Logging is an important aspect such that the devices should be logging to a central location so an operator could be able to derive the functionality of the entire network end to end as well as vulnerability management solution where a customer has a view of vulnerabilities in the network and can protect against them.

9. CONCLUSIONS AND FUTURE WORK

This thesis highlights the diverse definitions of network slicing between several SDOs aiming to provide a thorough definition of a baseline 5G platform for network slicing. Additionally, a comparison is being performed of three proof-of-concept network slice architectural approaches coming inherently with NFV and SDN principles in terms of practical implementation, technology adoption and deployment strategy. However, the design of a comprehensive solution to provide 5G mobile network services to vertical industries requires additional components to deploy and manage the slices providing the vertical services. Inspired by the third "Multiple owners or Multiple Tenants" model for the management and orchestration of network slices an OSS/BSS prototype called OpenSlice is recommended as representing an implementation from the industry. OpenSlice has the evolving platform to manage and control the NFV/SDN based complicated infrastructure and is suitable for network slicing. Network slicing is a very rich research area. How to create and manage the resources for that, what resources are to be sliced, and how to be integrated with other networks and finally how to manage their life cycle efficiently are some emerging areas of research. The last part of this thesis provides future challenges and research directions related to 5G network slicing. Also, a future work initiative will be to investigate the application of the proposed solution in real-world 5G and beyond trials and evaluate the perceived quality of its performance for various vertical business applications. In addition, the integration of the OpenSlice in conjunction to a multi-vendor Packet Core will emerge new initiatives for more efficient testing frameworks and FCAPS solutions which starting to be demanding for evaluation purposes for success of network slicing. The success of adaptive network slicing in a continuously broadened deployment will further be enhanced by evaluating the perceived quality of its performance through optimally multi-vendor 5G systems by continuously evolving the testing frameworks, FCAPS and Logging Solutions and overstretching these networks with traffic flows based on many different QCI parameter sets.

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviation	Definition
3D	3 Dimensional
3G	3rd Generation
3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
5GC	5G Core
AKA	Authentication and Key Agreement
AM	Access and Mobility
AMF	Access and Mobility Management Function
AN	Access Network
API	Application Programming Interface
AR	Augmented Reality
ARPF	Authentication Credential Repository Function
AUSF	Authorization User Subscription Function
BBF	Broadband Forum
BSS	Business Support System
CFS	Customer Facing Service
CN	Core Network
COMS	Common Operation and Management on Network Slices
CPU	Central Processing Unit
CRUD	Create, Read, Update, and Delete
CSP	Customer Service Provider
CU	Central Unit
CU-C	Central Unit for Control Plane Traffic
CU-U	Central Unit for User Plane Traffic
DNN	Data Network Name
DNS	Domain Name System
DRX	Discontinuous Reception
DU	Distributed Unit
E2E	End to End
ECGI	E-UTRAN Cell Global Identifier
EIR	Equipment Identity Register
EMBB	Enhanced Mobile Broadband
EPC	Evolved Packet Core
EPS	Evolved Packet System
ETSI	European Telecommunications Standards Institute
FCAPS	Fault, Configuration, Accounting, Performance, Security
GNB	G Node B
GSM	Global System for Mobile Communications
GSMA	Global System for Mobile Communications Association
GST	Generic Slice Template

GTP	Generic Transport Protocol
GTPC	Generic Transport Protocol Control Plane
GUTI	Global Unique Temporary Identifier
HPLMN	Home Public Land Mobile Network
HTTP	Hypertext Transfer Protocol
HTTP2	Hypertext Transfer Protocol 2
IETF	Internet Engineering Task Force
IAAS	Infrastructure as a Service
IMSI	International Mobile Subscriber Identity
IOT	Internet of Things
IP	Internet Protocol
IT	Information Technology
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LCM	Life Cycle Management
LTE	Long Term Evolution
MAC	Media Access Control
MANO	Management and Orchestration
MCC	Mobile Country Code
MIMO	Multiple Input Multiple Output
MIOT	Massive Internet of Things
MME	Mobility Management Entity
MNC	Mobile Network Code
MPLS	Multiprotocol Label Switching
NAI	network address identifier
NAS	Non-Access Stratum
NEF	Network Exposure Function
NEST	Network Slice Type
NF	Network Function
NFV	Network Function Virtualization
NFVO	Network Function Virtualization Orchestrator
NG	Next Generation
NGAP	NG Application Protocol
NGRAN	NG Radio Access Network
NR	New Radio
NRCGI	New Radio Cell Global Identifier
NRF	Network Repository Function
NS	Network Slice
NSA	Non-Standalone
NSAAS	Network Slice as a Service
NSD	Network Service Descriptor
NSI	Network Slice Instance
NSMF	Network Slice Management Function
NSSAI	Network Slice Selection Assistance Information
NSSF	Network Slice Selection Function
NSSMF	Network Slice Subnet Management Function

OAUTH	Open Authorization
ONF	Open Network Foundation
ORAN	Open Radio Access Network
OSI	Open Systems Interconnection
OSS	Operations Support System
OTT	Over the Top
PAAS	Platform as a Service
PCF	Policy Control Function
PDPC	Packet Data Convergence Protocol
PDU	Protocol Data Unit
PEI	permanent Equipment Identifier
PFCP	Packet Forwarding and Control Protocol
PGW	Packet Gateway
PHY	Physical Layer
PLMN	Public Land Mobile Network
PNF	Physical Network Function
QCI	QoS Class Identifier
QOE	Quality of Experience
QOS	Quality of Service
RAN	Radio Access Network
RAT	Radio Access Type
RFS	Resource Facing Service
RIC	RAN Intelligent Controller
RLC	Radio Link Control
RRC	Radio Resource Control
RRM	Radio Resource Management
RT	Real Time
RU	Radio Unit
S1AP	S1 Application Protocol
SA	Standalone
SAAS	Software as a Service
SB	Service Blueprint
SBA	Service Based Architecture
SBCAP	Cell Broadcast Centre interfaces with the Evolved Packet Core
SC	Service Catalog
SCSAP	SCs Application Protocol
SCTP	Stream Control Transmission Protocol
SD	Service Differentiator
SDN	Software Defined Network
SDNO	Software Defined Network Orchestrator
SDO	Standards Definition Organizations
SEAF	Security Anchor Function
SEPP	Security Entity Proxy
SGW	Serving Gateway
SIM	Subscribed Identity Module
SLA	Service Level Agreement

SLSAP	SLs Application Protocol
SM	Session Management
SMF	Session Management Function
SO	Service Orchestrator
SOM	Service Order Management
SSC	Session and Service Continuity
SST	Slice Service Type
SUCI	Subscriber Unique Concealed Identity
SUPI	Subscriber Unique Permanent Identity
TA	Tracking Area
TAAS	Testing as a Service
TAP	Test Automation Project
TC	Test Case
TCP	Transmission Control Protocol
TE	Traffic Engineering
TLS	Transport Layer Security
TM	Tele Management
TMF	Tele Management Forum
TMSI	Temporary Mobile Subscriber Identity
TS	Technical Specification
TTI	Transmission Time Interval
UDM	Unified Data Mediation Function
UE	User Equipment
UI	User Interface
UICC	Union for International Cancer Control
UPF	User Plane Function
URLLC	Ultra-Low Latency Communications
VIM	Virtualized Infrastructure Manager
VINNI	Verticals Innovation Infrastructure
VLAN	Virtual Local Area Connection
VM	Virtual Machine
VNF	Virtualized Network Function
VNFD	Virtualized Network Function Descriptor
VNFM	Virtual Network Functions Manager
VPLMN	Visited Public Land Mobile Network
VR	Virtual Reality

APPENDIX I

I.1 Networking Servicing Terms

I.1.1 Communication Network

A communication network (or simply a “network”) refers to a communication between different devices or Nodes or Functions with the aim to provide connection between Endpoints subject to defined underlying rules and mechanisms to achieve routing with certain protocols. These Endpoint devices are known as User Equipment. The UEs are configured to communicate with the network either via a fixed line connection or via a wireless connection. The UE includes a variety of connected devices including UEs as defined by the 3rd Generation partnership project (3GPP), mobile devices (e.g., wireless handsets) and other connected devices, including Machine-to-Machine (M2M) devices (also referred to as Machine Type Communications (MTC) devices) or small stationary or mobile devices like robots or sensors eg.in an industrial environment. So, the network should be able to provide communication services either in a stationary UE or a Mobile UE by providing the same quality of Experience and by using a unified underlying architecture. Following this rational a network may include, for instance, at least one of a radio access portions which interfaces directly with UEs via radio access and a fixed line portion which interfaces directly with UEs via fixed line access, in combination with a unified backhaul portion which connects different network devices of the network together. So, a 5G network is required to be used in many different use cases from telecom operators providing many different types of services, to Non private networks like industrial environments. Such a network should be able to be reconfigured to suit various needs and thus comprises various virtualized components. Virtualization is the key technology allowing the network to support network slicing to create different sub-networks with characteristics suited for the needs of the traffic they are designed to support. The network may include a number of computing hardware resources that provide processors and/or allocated processing elements, memory, and storage to support functions executing on the network, as well as a variety of different network connectivity options connecting the computing resources to each other and making it possible to provide service to User Equipment.

I.1.2 Service

A service for the most part relates to a software that performs at least one functions and provides APIs to applications or other services of the equivalent or various layers to utilize said functions and returns at least one outcome. Services can be joined with different services, or brought in a specific serialized way, to create another service. Accessing a service may involve communication between multiple endpoints that are connected to the network. A service may be provided by the network operator or may be provided by a network customer such as a business, utility, government, or other organization. Examples of services include, but are not limited to, providing audio and/or video content to stream or download to an endpoint such as a UE, storage and/or processing of data from an endpoint such as a UE, UE-to-UE messaging services, machine-to-machine communications such as utility meter reporting, remote data storage, and/or remote computing services.

I.1.3 Network Slicing Terms

A network slice generally corresponds to a set of network resources which have been allocated to support at least one specific service on the network. A Network Slice is a complete logical network including Radio Access Network (RAN) and Core Network (CN). It provides telecommunication services and network capabilities, which may vary (or not) from slice to slice. Distinct RAN and Core Network Slices will exist. A UE may access various Network Slices at the same time through a solitary RAN. The process that a UE initiates to a service and/or network slice starts in known as the Registration Procedure. Registration is initiated through a radio access node RAN that is currently providing connection between the UE and the network.

I.1.4 Network Resource Terms

I.1.4.1 Network Entity

A network entity generally refers to a network node, or a combination of network nodes, that is operative to provide specified services on the network. Is defined as a manageable logical entity uniting one or more network devices. This allows distributed devices to be managed in a unified way using one management system. It also means a facility or equipment used in the provision of a communication service. A network entity comprises physical components which may be dedicated physical components, or the network entity may be allocated use of the physical components of another device. A network entity may be associated with multiple physical components that may be located either in one location or may be distributed across multiple locations.

I.1.4.2 Network Function

A network function comprises a service that may be provided by a network entity. A Network function can be implemented as a network node on a dedicated hardware or as a virtualized software function. It includes but is not limited to network nodes functionality, e.g., session management, mobility management, switching, routing functions, which has defined functional behavior and interfaces. Data, Control, Management, Orchestration planes functions are Network Functions.

REFERENCES

[1] T. Specification and G. Services, “3Gpp Tr 21.916 Release 16 Description,” 2020. https://www.3gpp.org/release-16.

[2] “3GPP Release 17.” https://www.3gpp.org/release-17.

[3] T. Specification and G. 3GPP Services, “3Gpp TS 23.501 System Architecture for the 5G System,” vol. 16.0.2, no. Release 16. p. 308, 2019, [Online]. Available: http://www.3gpp.org.

[4] B. Bertenyi, R. Burbidge, G. Masini, S. Sirotkin, and Y. Gao, “NG Radio Access Network (NG-RAN),” *J. ICT Stand.*, vol. 6, no. 1, pp. 59–76, 2018, doi: 10.13052/jicts2245-800x.614.

[5] T. Specification, G. Core, N. Function, and R. Services, “3Gpp Ts 29.510,” vol. 0, no. Release 15, 2019, [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3345.

[6] G. Mayer, “RESTful APIs for the 5G ServiceBased Architecture,” *J. ICT Stand.*, vol. 6, no. 1, pp. 101–116, 2018, doi: 10.13052/jicts2245-800x.617.

[7] G. Liu, Y. Huang, Z. Chen, L. Liu, Q. Wang, and N. Li, “5G Deployment: Standalone vs. Non-Standalone from the Operator Perspective,” *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 83–89, 2020, doi: 10.1109/MCOM.001.2000230.

[8] A. Morgado, K. M. S. Huq, S. Mumtaz, and J. Rodriguez, “A survey of 5G technologies: regulatory, standardization and industrial perspectives,” *Digit. Commun. Networks*, vol. 4, no. 2, pp. 87–97, 2018, doi: https://doi.org/10.1016/j.dcan.2017.09.010.

[9] S. Köksal, “Evolution of Core Network(3G vs. 4G vs. 5G).” 2019, [Online]. Available: https://medium.com/@sarpkoksal/core-network-evolution-3g-vs-4g-vs-5g-7738267503c7.

[10] Huawei, “5G Network Architecture A High-Level Perspective, White Paper,” *Huawei Technol. Co., Ltd*, no. July, 2016, [Online]. Available: https://www.huawei.com/minisite/5g/img/5G_Network_Architecture_A_High-Level_Perspective_en.pdf.

[11] T. Specification and G. Services, “3Gpp Tr 28.801,” vol. 0, no. Release 15, pp. 1–75, 2018.

[12] I. D. Telefonica, “COMS Architecture,” no. c, pp. 1–12, 2018, [Online]. Available: https://datatracker.ietf.org/doc/draft-geng-coms-architecture/.

[13] Huawei Technologies, “The Use Cases of Common Operation and Management of Network Slicing,” 2018, [Online]. Available: https://tools.ietf.org/html/draft-qiang-coms-use-cases-00.

[14] S. Bryant and J. Dong, “Network Slicing Architecture,” no. April, pp. 1–8, 2018, [Online]. Available: https://tools.ietf.org/id/draft-geng-netslices-architecture-01.html.

[15] S. Chiosi, Margaret *et al.*, “Network Functions Virtualisation (NFV),” *ETSI white Pap.*, no. 1, pp. 1–20, 2015, [Online]. Available: https://www.etsi.org/deliver/etsi_gr/NFV-EVE/001_099/012/03.01.01_60/gr_NFV-EVE012v030101p.pdf.

[16] GSM Alliance, “An Introduction to Network Slicing,” *White Pap.*, 2017, [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf.

[17] R. Trivisonno, X. An, and Q. Wei, “Network slicing for 5G systems: A review from an architecture and standardization perspective,” in *2017 IEEE Conference on Standards for Communications and Networking (CSCN)*, Sep. 2017, pp. 36–41, doi: 10.1109/CSCN.2017.8088595.

[18] ETSI, “5G; Procedures for the 5G System,” 2018, [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123502/16.07.01_60/ts_123502v160701p.pdf.

[19] G. P. Project, T. Specification, and G. Services, “3Gpp Ts 33.501,” vol. 0, no. Release 16, 2020, [Online]. Available: https://www.etsi.org/deliver/etsi_ts/133500_133599/133501/15.04.00_60/ts_133501v150400p.pdf.

[20] D. Basin, J. Dreier, L. Hirschi, S. Radomirović, R. Sasse, and V. Stettler, “A Formal Analysis of 5G Authentication,” Jun. 2018, doi: 10.1145/3243734.3243846.

[21] F. H. P. Fitzek, *Computing in Communication Networks From Theory to Practice*. 2019.

[22] M. A. Habibi, B. Han, and H. D. Schotten, “Network slicing in 5G mobile communication:

Architecture, profit modeling, and challenges," *arXiv*, 2017.

[23] T. Tovinger, "Management, Orchestration and Charging for 5G networks," *3Gpp*, 2018. http://www.3gpp.org/NEWS-EVENTS/3GPP-NEWS/1951-SA5_5G.

[24] A. A. Barakabitze, A. Ahmad, R. Mijumbi, and A. Hines, "5G network slicing using SDN and NFV: A survey of taxonomy, architectures and future challenges," *Comput. Networks*, vol. 167, 2020, doi: 10.1016/j.comnet.2019.106984.

[25] H. Flinck, C. Sartori, A. Andrianov, C. Mannweiler, N. Sprecher, and Nokia, "Network Slicing Management and Orchestration draft-flinck-slicing-management," *Internet Eng. Task Force*, p. 26, 2017, [Online]. Available: https://tools.ietf.org/id/draft-flinck-slicing-management-00.html.

[26] GSMA, "Network Slicing Use Cases Requirements," *Futur. Networks Program.*, vol. 1, no. April, p. 54, 2018, [Online]. Available: https://www.gsma.com/futurenetworks/wp-content/uploads/2018/03/Network-Slicing-Use-Cases-Requirements-Wrapper.pdf.

[27] "5G Victory: Field trials methodology and guidelines," 2020, [Online]. Available: https://www.5g-victori-project.eu/wp-content/uploads/2020/10/2020-09-25-5G-VICTORI_D4.1_v1.0_Website_Version.pdf.

[28] C. Tranoris, "Openslice: An opensource OSS for Delivering Network Slice as a Service," [Online]. Available: https://orcid.org/0000-0002-3433-037X.

[29] J. Ordonez-Lucena, C. Tranoris, and J. Rodrigues, "Modeling network slice as a service in a multi-vendor 5g experimentation ecosystem," *2020 IEEE Int. Conf. Commun. Work. ICC Work. 2020 - Proc.*, 2020, doi: 10.1109/ICCWorkshops49005.2020.9145225.

[30] A. Gavras *et al.*, *Onboarding Vertical Applications on 5G-VINNI Facility*. 2020.

[31] J. Ordóñez, C. Tranoris, J. Rodrigues, and L. Contreras, *Cross-domain Slice Orchestration for Advanced Vertical Trials in a Multi-Vendor 5G Facility*. 2020.

[32] C. Tranoris and S. Denazis, *A Workflow for Onboarding Verticals on 5G/NFV Experimental Network Facility*. 2020.

[33] T. S.A., "PoC Project: Automated network slice scaling in multi-site environments," 2020.

[34] M. Chahbar, G. Diaz, A. Dandoush, C. Cerin, and K. Ghoumid, "A Comprehensive Survey on the E2E 5G Network Slicing Model," *IEEE Trans. Netw. Serv. Manag.*, vol. 18, no. 1, pp. 49–62, 2021, doi: 10.1109/TNSM.2020.3044626.

[35] IETF, "Network Slice Provision Models," 2019, [Online]. Available: https://www.ietf.org/id/draft-homma-slice-provision-models-02.txt.

[36] K. Qu, W. Zhuang, Q. Ye, X. Shen, X. Li, and J. Rao, "Traffic Engineering for Service-Oriented 5G Networks with SDN-NFV Integration," *IEEE Netw.*, vol. 34, no. 4, pp. 234–241, Jul. 2020, doi: 10.1109/MNET.001.1900508.

[37] Eugina Jordan, "Open RAN 101–Role of RAN Intelligent Controller: Why, what, when, how? (Reader Forum)," [Online]. Available: https://www.rcrwireless.com/20200730/opinion/readerforum/open-ran-101-role-of-ran-intelligent-controller-why-what-when-how-reader-forum.

[38] ONF & 3GPP, "What is RAN Disaggregation?," 2021. https://moniem-tech.com/2021/02/16/what-is-ran-disaggregation/.

[39] Huawei, "5G Mec IP Network White Paper," 2020, [Online]. Available: https://carrier.huawei.com/~/media/CNBGV2/download/program/5G-MEC-IP-Network-White-Paper-en-v2.pdf.

[40] J. Bright, "Driving New Business Opportunities with Multi-Access Edge Computing and 5G," 2019, [Online]. Available: https://carrier.huawei.com/~/media/CNBGV2/download/products/core/OVUM-en.pdf.

[41] R. K. Srinivasa, N. Kumar, S. Maheshwari, C. Bharathi, and A. Kumar, *Minimizing Latency for 5G Multimedia and V2X Applications using Mobile Edge Computing*. 2019.

[42] Z. Kotulski *et al.*, "On end-to-end approach for slice isolation in 5G networks. Fundamental challenges," *Proc. 2017 Fed. Conf. Comput. Sci. Inf. Syst. FedCSIS 2017*, pp. 783–792, 2017, doi: 10.15439/2017F228.