Resource Sharing for a 5G Multi-tenant and Multi-service Architecture

Marcos Rates Crippa*, Paul Arnold†, Vasilis Friderikos§, Borislava Gajic‡, Carmen Guerrero*, Oliver Holland¶, Ignacio Labrador†, Vincenzo Sciancalepore**, Dirk von Hugo†, Stan Wong‡, Faqir Zarrar Yousaf**, Bessem Sayadi§

* University of Kaiserslautern † Deutsche Telekom Technology Innovation § Universidad Carlos III de Madrid ¶ Nokia Bell–Labs ‡ King’s College London † Atos Research and Innovation ** NEC Europe Ltd.

Abstract—As the next generation networks (5G) move into a direction of virtualization and softwarization, using technologies like Software-Defined Networking (SDN) and Network Function Virtualization (NFV), one of the main benefits is the possibility of network resource sharing among different tenants and service providers. This paper focuses on how the architecture of the 5G NORMA project will tackle the issue of resource sharing. Two main areas will be explored: how to achieve efficient resource sharing among network slices, and specific solutions used by the architecture to allow for resource sharing.

Index Terms—NFV; SDN; Orchestration; 5G networks; Resource Sharing; MANO; Network Slicing

I. INTRODUCTION

The future 5G networks are expected to rely on softwarization and virtualization of network elements and functions to support flexible and heterogeneous deployment of different services. This deployment could be on the same infrastructure, with services as diverse as Internet of Things (IoT) and Vehicular Networking expected to efficiently share the network resources. A re-engineering of the current 3GPP architecture is necessary, in order to add the flexibility to support diverse services with diverge requirements. One aspect of this re-engineering is to design efficient resource sharing among different tenants, operators and network slices.

Network slicing [1] is one innovation 5G NORMA [2] uses to enable the future 5G networks. A network slice can be broadly defined as an end-to-end logically isolated virtual network that includes access, transport and core network functions. In general, these network functions can also be shared between different slices based on pre-defined policies and business criteria. An abstraction of different hardware infrastructures into a logical virtual network is necessary to allow for this sharing. Various virtual network functions (VNFs), created by the decomposition of network elements, will operate in this logical network. Deploying separate hardware infrastructures for each service would not be a cost effective way for efficient resource sharing among multiple tenants. On the other hand, network slices can be instantiated as a mix of dedicated and shared resources (such as transmission points, radio resources, transport and fronthaul capacity, potentially IT resources, for instance), allowing for multi-service and multi-tenant networks.

This paper focuses on the 5G NORMA approach to shared resource management between network slices in its 5G architecture, as well as technical solutions aiming at providing the architecture with tools for resource sharing. Initially, an overview of the 5G NORMA architecture is given. Following that, the challenges of sharing resources across network slices are described. The different proposed technical solutions addressing this issue in 5G NORMA are listed and detailed next. Finally, a brief discussion of the relationship between the SDM-C and SDM-X components of the 5G NORMA architecture concludes the paper.

II. THE 5G NORMA ARCHITECTURE

To build a flexible and adaptable mobile network architecture capable of supporting a wide variety of services and their respective requirements, 5G NORMA has introduced a novel paradigm: a network-of-functions-based architecture. This novel paradigm breaks the design principle followed by current network architectures, which are built around entities rather than functions. Our revolutionary approach builds on new technologies, such as Software-Defined Networking (SDN) and Network Function Virtualization (NFV), in conjunction with novel concepts such as the network slicing and multi-tenancy. More details on the relationship between slices and tenants can be found in [13]. This section presents a brief description of the architectural concepts of 5G NORMA related to its various controllers’ design and specification, mobility management, and network slice management. Three different controllers are considered in the architecture, namely the Software Defined Mobile Network Controller (SDM-C), the Software Defined Mobile Network Coordinator (SDM-X), and the Software Defined Mobile Network Orchestrator (SDM-O) as a part of the MANO framework. We refer the interested reader to [2] for further details.

Figure 1 presents a functional overview of the 5G NORMA architecture. These functional blocks will be referred in the subsequent sections.

A. SDM-C

The SDM-C applies the same principles of the current Software-Defined Networking (SDN) to wireless functionality beyond routing. Indeed, the benefits of this technology when
applied to wireless networks are even more significant than for wired networks, as the control functionality of wireless networks include many additional and more complex functions than just routing. This includes time-critical functions (such as scheduling control or self-organizing networks capabilities) and other less time critical (such as Radio Resource Control, power control and handover decision and execution). With SDM-C, all these functions are implemented in a centralized functions, decoupling the network intelligence from the underlying hardware infrastructure by a programmable central control, which provides very important benefits for the flexible operation in the edge of the wireless network [2].

B. SDM-O and MANO framework

The SDM-O, which is part of the Management & Orchestration layer of the 5G NORMA architecture, is an extended version of the ETSI NFV MANO’s NFVO functional block with the added feature of an Inter-slice Resource Broker [3] to enable and manage end-to-end network slices. Another aspect that differentiates an SDM-O from the ETSI NFV MANO is that the Inter-slice Resource Broker enables the SDM-O to support multi-tenancy. Each tenant has its own NFV-MANO stack for the management and orchestration of its respective network service (NS), which can be considered as a service slice, and the SDM-O via the Inter-slice Resource Broker manages and orchestrates between these different tenants. This arrangement enhances the operational/functional features of the base NFV MANO system towards providing a harmonized support of 5G requirement for agile management of multi-tenants and their resources/services in a single and/or multiple administrative domains. The SDM-O also interfaces with the Service Layer via the Service Management functional block, which transforms the consumer-facing service descriptions into resource-facing service descriptions and vice-versa. Furthermore, the SDM-O interfaces with the SDM-C in the control layer in order for the SDM-C to directly trigger the re-orchestration request on the SDM-O in case the SLA targets cannot be met directly by the SDM-C.

C. SDM-X

The SDM-X is responsible for the control of resources and network functions shared among network slices, such as the spectrum and the hardware infrastructure, among others. Thus, to operate multiple network slices in the same infrastructure efficiently, a common entity (SDM-X) is introduced in the architecture. The problem of resource sharing among network slicing (i.e., two slices using the same MAC network function) is the main focus of this paper, and some selected algorithms that make use of this controller are described in Section IV.

D. Mobility Management

5G services and slices exhibit different demands for mobility support in terms of e.g., terminal speed, session continuation requirements, and of stability of the endpoint address. A corresponding suitable mobility management (MM) schemes can differ in many ways, e.g., requiring special handover policies and settings in the RAN, flexible mobility anchoring, adaptive gateway relocation rules, or customized network elements (e.g., local gateways or gateways with specific mobility support).

The management of user mobility considering the instantaneous traffic demands and required KPIs is a thorough process that in 5G NORMA involves re-orchestration by SDM-O and re-configuration by SDM-C to achieve an optimized network performance. Mobility Management in 5G NORMA is managed as SDM-C application (MM App). Selection of appropriate MM algorithms, type of parameters and values to be configured within a service-aware MM App (i.e. describing a VNF template for a specific type of MM scheme) needs to be done according to the specific network slice requirements. The MM App might also be shared among different network slices, further details about are given in Section IV.

E. QoS/QoE Monitoring and Mapping

In order to optimize the network resource utilisation, continuous monitoring of QoS and QoE for each service flow is required. QoS and QoE are computed by mapping objective network parameters (e.g., jitter, throughput, latency) and certain subjective influence factors (e.g., environment, user age, terminal type) that could affect the users experience. Such objective network parameters (KPIs) are obtained by network performance measurements on different network elements (e.g. mobile terminals, Radio Remote Head) and are fed into a QoS/QoE Monitoring and Mapping module of 5G NORMA. The QoS/QoE Monitoring and Mapping module supports a fully dynamic, context-aware QoE/QoS management that is able to regard the applications along with their QoS/QoE requirements and jointly with SDM-O/C/X adapt the end-to-end resource allocation and the data plane services accordingly.

F. Intra-slice Management

Each network slice has an SDM-C, responsible for managing the network slice resources and building the paths to join the network functions taking into account the received requirements and constraints which are being gathered by the QoS/QoE Monitoring and Mapping module. The SDM-C, based on slice performance reports received by QoS/QoE Monitoring and Mapping module, may adjust the network slice configuration either by reconfiguring some of the VNFs in a network slice or by reconfiguring data paths in a SDN-like style. The QoS/QoE module along with the SDM-C constitute
the intra-slice management. If the requirements cannot be met by aforementioned reconfigurations of VNFs or data paths the SDM-O can perform a slice reshaping e.g., by adding more resources to the given network slice.

G. Inter-slice Management

The SDM-O has a complete knowledge of the network managing the resources needed by all the slices of all tenants. This enables the SDM-O to perform the required optimal configuration in order to adjust the amount of used resources. While the SDM-C directly interfaces with dedicated NFs, the SDM-X controls shared NFs. Together with the SDM-O the SDM-X constitutes the inter-slice management. The inter-slice management and orchestration is a key-feature of the novel 5G NORMA architecture as it fosters and supports multi-service and multi-tenancy systems.

III. NETWORK SLICE MANAGEMENT AND RESOURCE SHARING

Since network slices are instantiated from a common underlying infrastructure an appropriate assignment of hardware/software/radio resources and functions among them is necessary. Initially, this is performed enabling the context-aware adaptation and allocation of the different NFs for each slice in the front-end (e.g., radio site) and in the edge or central clouds thanks to the possibility of using fine-grain NFs (de)composition. But also, to support the different slices instantiation assigning the necessary resources to each slice, 5G NORMA provides a specific Management and Orchestration (MANO) layer to manage the creation of each slice assigning the necessary network services for each case. This layer takes as input the specific SLA parameters for each tenant (generated from the Service Management block), and generates a slice description blueprint describing the different NFs in the slice, NFs topology (NFs forwarding graph), QoS/QoE parameters and the specific instructions for the slice deployment, configuration, monitoring and rules for its orchestration.

A central component in this MANO layer is the SDM-O, which manages the resources needed by all the slices of all tenants; basically it exposes interfaces to process network slices creation, operation, and termination requests to the Service Management block, which in turn serves the infrastructure provider Operations and Business Support Systems (OSS/BSS). The SDM-O performs two main functions: (a) a generic resource provisioning and (b) resource pooling/reservation. Resource provisioning is performed to calculate and provide the right amount of virtualized infrastructure resources (e.g., computing or storage) according the agreed SLA between the tenant and the Infrastructure Provider; on the other hand, resources pooling and reservation refers to the process of combining and jointly managing large-scale resources to simultaneously serve multiple tenants. These functions make it possible to perform the optimal configuration in order to adjust the amount of used resources and, hence, get an efficient use of the network.

By executing these two functions the SDM-O creates each network slice instance fulfilling the required parameters and using the available infrastructure resources. So, the role of the SDM-O is to map the requirements defined in the SLA by instantiating and configuring the necessary Network Functions into dedicated network slices, i.e., it performs the Inter-Slice orchestration function, having an overall view of all the slices across all the resource domains (radio, edge and central clouds); on the other hand, Intra-Slice orchestration can be performed internally for each slice within each slice specific scope; this is performed by a slice specific NFVO block. So, the SDM-O decides about the optimal sharing of resources (both computational and networking-related) and how they have to be assigned to the different network slices, while the NFVO performs orchestration at slice level. For a better control of the SLA enforcement the SDM-O is also able to process real-time triggers from the different slices requesting certain re-orchestration operations (e.g., scale up/down requests on certain NFs).

Besides the SDM-O in the MANO layer, the SDM-C and SDM-X are used to control resources to each slice. The first one enables the control of dedicated network functions, allowing a flexible network management and operation of resources within the network slices and controlling both: data plane and control plane nodes. On the other hand, the SDM-X receives orchestration information from the SDM-O and processes it in order to decide whether it is necessary or not to modify the network slices shared resources; it also coordinates among multiple SDM-C instances and the associated network slices.

SDM-C and SDM-X control Physical Network Functions (PNFs) and the software applications in VNFs, but they do not control VNFs underlying NFVI resources. For VNFs this implies that SDM-C/X only control the software running into virtual machines. The virtualization container of the VNF is controlled by the NFV MANO entities (e.g., VIM and VNFM). This ensures a consequent split between management/control of mobile network functions on the one hand and NFVI resources on the other. For PNFs, which exhibit a tight coupling of hardware and software, SDM-C and SDM-X control the entire HW/SW system of the network functions.

IV. SHARED RESOURCE MANAGEMENT FOR 5G NETWORKS

A number of solutions have been proposed under the 5G NORMA architecture to achieve resource sharing between different network slices.

A. ICIC Schemes

As described in Section II-C, the SDM-X is in charge of controlling shared radio resources. To meet slice specific service requirements in cases when BSs need to serve multiple slices with more than one service data flow each, it is crucial to react dynamically on critical interference situations in the network. The SDM-X needs the opportunity to optimize temporary appearing critical interference constellations in local parts of the mobile network. It should influence synchronous inter cell interference coordination (ICIC) schemes by dynamically adapting BS clusters (e.g. for Joint Transmission (JT), Coordinated Multipoint (CoMP) or coordinated beamforming) or it can adapt and de-/active asynchronous ICIC schemes (e.g.,
frequency reuse schemes, Carrier Aggregation (CA) based ICIC, enhanced (e)ICIC). It might by even possible to de-/activate the Medium Access Control (MAC) scheduler, if alternative schemes were orchestrated by the SDM-O. Fig. 2 shows the principle idea. A database with RRM applications, such as ICIC and scheduling schemes is defined and some of them will be placed to the physical nodes of the mobile network during the orchestration process and executed as NBI application through SDM-X. The SDM-X takes care of the control of the RRM schemes during life cycle management. Fig. 3 illustrates a simplified message sequence chart based on [12]. The procedure is described in the following:

After the SDM-O has set up at least one network slice on the physical infrastructure, service policies regarding QoS and SLA requirements for each slice are transmitted to the SDM-C based monitoring entity as well as to the SDM-X. In addition, it is necessary to provide further information to the SDM-X. It needs to know which VNFs are located to which physical nodes within the infrastructure. After successfully received the VNF mapping table the SDM-X takes over the control to influence the service chaining regarding the radio scheduling decision among shared resources of the MAC scheduler. As an alternative, the service chain adaptation could be included natively in the application in the sense if the app detects some trigger it will send new information via SDM-X to the VNFs. The SDM-X is in charge of de-/activating asynchronous ICIC schemes and defines dynamically radio resource and BS cluster maps for synchronous scheduling schemes based on the feedback of the slice specific QoS/QoE monitoring entities (Section II-E) Once, multiple SLA indicators, derived from the QoS/QoE Monitoring module, such as a total amount of average throughput in a certain time period or a maximum delay probability are received the SDM-X checks whether a previously defined threshold violation occurs (step 1). Based on the analysis of slice specific threshold violation the SDM-X derives a decision to switch the ICIC, scheduling strategy or to change a BS cluster. This can be repeated several times to improve the SLA of a threatened slice until a more critical threshold is violated. After adapting the service chaining according to the ICIC scheme or the scheduling metric, the SDM-X considers the monitoring information again (step 2). If the adaptation of the scheduling, ICIC scheme or the possible change of a BS cluster did not result into a performance improvement for currently instantiated slices and the SLA requirements cannot be fulfilled, the SDM-X needs to send a modification request to the SDM-O, which has to re-orchestrate (step 3) the slices based on different service chain templates.

Fig. 2: SDM-X control of ICIC/scheduling schemes

**B. Network Slice Resource Brokering**

When multiple slices are instantiated, slice brokering operations on Radio Access Network (RAN) facilities are required to check the resource availability before allocating new network slice request resources, and prevent the system from violating service agreements for previous allocated slice requests. This is performed through a two-steps procedure: the former is a slice monitoring phase, the latter is a slice traffic scheduling phase [6]. In the monitoring phase, the SDM-X dynamically manages and assigns network resources, e.g., physical resource blocks (PRBs), to different slices by means of scheduling policies applied to conventional traffic schedulers. The SDM-X can easily monitor given performance figures, such as utilization rate over fixed time windows, peak or average rates in terms of bit per seconds and so further. Such statistics are used as input for advanced forecasting algorithms, which may prevent SDM-X from allocating resources to future slice requests in future time periods. On the other side, accurate predictions might help the SDM-X to efficiently allocate shared RAN resources. This concept leverages on the potential gap between the amount of resources asked in the network slice request and the real utilization of such resources. However, advanced schemes are needed to provide the service quality guarantees agreed during the network slice negotiation with the SDM-O, as mentioned in Section II-B. During the scheduling phase, the SDM-X efficiently decides the resource block set to assign to each network slice. This operation is performed through a network resources mask, an abstraction of real network resources to be fully used by each network.
slice. The network resources mask is dynamically changed by the SDM-X to continuously fulfill the slice SLAs. In case of prediction failures or unexpected traffic bursts, the SDM-X might trigger an alert message to the SDM-O, and, in case of congestion, it can reject incoming network slice requests.

C. Spectrum Sharing

Spectrum sharing will be vital in the context of 5G capacity demands, particularly those demands at lower frequencies for which far less available spectrum bandwidth exists. Moreover, vast increases in lower-frequency spectrum access will be necessary to underpin the service coverage unreliability of higher-frequency mm-wave access in 5G.

5G NORMA has noted important recent developments in (broadly-speaking) geolocation database (GDB)-based spectrum sharing, such as the Citizens Broadband Radio Service (CBRS) in the US, Licensed-Shared Access (LSA) concepts in Europe [9], and TV White Space (TVWS) in the US, UK, Singapore, Japan and numerous other administrations [9], [10]. It has incorporated the GDB concept into its SDM capabilities in order to manage spectrum among different incumbents, or among incumbents and new entrants, or even in contexts such as contended access to unlicensed spectrum.

The GDB broadly works through the handling of resource requests, in which devices, base stations and other elements convey their locations and technical capabilities. The GDB responds indicating available resources, based on which the aforementioned elements send information on their resource usage choices back to the GDB. Periodic resource usage reconfirmation requests or urgent relinquishment messages (e.g., in the case of an incumbent reappearing) are also supported. 5G NORMA proposes that the GDB is deployed in a distributed fashion, using concepts such as a distributed ledger to appropriately filter allowed resource usages at different levels—with the distributed GDB elements at the lowest level being very close to the spectrum users and therefore able to address resource requests with a very low latency. This will assist applications such as URLLC.

It is noted that the GDB in 5G NORMA is also prominently intended to serve other purposes, such as geographically-managing computational resources based on demand, and managing rendezvous in heterogeneous networking contexts, among others.

D. VNF placement consideration

In addition to spectrum also the question of sharing or exclusive usage of other network assets such as VNFs is a major issue in design of 5G systems. Operation of different slices for different tenants has to grant isolation between the slices, which is greatly eased by employing dedicated network functions running on separate resources. Sharing of common infrastructure and joint NFs across multiple slices would increase the efficiency in terms of resource utilization, but require more effort in terms of security and guaranteed operational isolation. Typically distributed PNFs are the radio access points jointly used for all services offered by an operator via the same technology. On the other hand an example for a centrally deployed common VNF is a user and subscription data base (comparable to the HSS in 3GPP) shared by all slices within the administrative domain of one operator Another access agnostic Core Network (CN) based Control Plane (CP) VNF is Mobility Management (MM) covering multiple subfunctions e.g. for anchoring, forwarding, and location management. The degree of sharing strongly depends on the slice context and characteristics, especially when multiple slices jointly use a common physical infrastructure for access and/or transport. The MM in CN can be deployed in different flavors of sharing: in case of a completely dedicated MM all subfunction instances are implemented separately per slice, while a partially shared function across slices would re-use e.g. a common anchoring DP function (e.g. a common MAG in PMIP) but separate location managers (for privacy preservation) or mobility decision entities according to diverging requirements. The chosen level influences also the exact implementation e.g. in terms of assigned computing resources and topological location of the MM. Multiple vehicular slices instantiated for different car manufacturers might be deployed in a fully shared manner re-using same anchor and forwarding entities (e.g. at Mobile Edge Clouds) controlled by a common MM-App via SDM-C - as shown in Figure 1 - since the slices are characterised by similar performance demands. Advantages and draw-backs of the three options for mobility management implementation based on level of CN sharing among network slices are sketched below:

1) Dedicated MM in a dedicated CN for each network slice facilitates slice separation in a straight forward way (e.g. allowing operation in different physical nodes) thus optimally meeting the service requirements
2) Hybrid MM in CN comprises both components shared among networks slices and components assigned to dedicated slices. E.g., MM execution and enforcement is shared between the network slices, while different policies and session management are provided and implemented per separate CN instances of network slices. Such an approach could require considerable effort for cooperation between SDM-C/-X/ and -O.
3) Common MM is deployed completely as a shared VNF with all subfunctions jointly controlling multiple slices enabling maximum resource efficiency.

Operation of dedicated VNFs per network slice might also result in additional parallel signaling within access, transport, and core network domains while on the other hand, VNFs designed as common entity across slices may lack perfect adaptation to specific demands, e.g. in case the slices experience different traffic load. As elaborated above the best choice of different deployment options depends on the required functionality and the slice specific constraints. So the decision has to carefully consider and analyse the expected specific situation and environment for which the slice is configured and additionally also whether a re-configuration during life cycle should be enabled.

E. SDM-C/X Authentication

5G NORMA has developed a hierarchical and distributed authentication, authorization and accounting (AAA) approach
at the edge cloud. This approach is named Virtualized-Authentication, Authorization and Accounting (V-AAA) [13], which is based on ETSI NFV [11] tokenization techniques to protect virtual network entities and tenants. Basically, the V-AAA also uses tokenization techniques for securely identifying, accessing, terminating, provisioning and deploying the virtual network entities, e.g., SDM-C and SDM-X, and services, e.g., network slices, via a provisioning platform. Furthermore, when tenant requires to create, manipulate or terminate network resources in SDM-C and SDM-X, it also uses the same sequence to access network or service resources from requesting to granting of the authorization, and from gaining to using the identity and access token to a specific network entity and resources in a particular period of time.

![Fig. 4: Basic SDM-C and SDM-X authentication sequences.](image)

**V. RELATIONSHIP BETWEEN SDM-X AND SDM-C**

SDM-C and SDM-X have been considered so far as being logically independent. However, there can be instances that their roles can be blurred depending on the algorithmic aspect on network optimization in use. Below we outline different possible interactions between those two elements.

1) **Independent Operation:** In this case a specific resource allocation per slice is already being defined and as a result any type of resources related to shared one can be provided by the SDM-X entity in a rather independent manner (in terms of the SDM-C). The allocation of shared resources does not impact the allocation from the SDM-C. This type of operation can be considered as the nominal one, i.e., these two entities can run their own algorithms with the requirement of an interface between them so that specific information to be exchanged regarding when to trigger different decision making algorithms. But, as mentioned below some other modes of operation can be envisioned that deserve further study and is left as an item of future research.

2) **Loose Coupling:** This case refer to the situation where resources per tenant are predefined as well as the amount of shared resources are also predefined but without specifying which resources from the available resource pool (for example Physical Resource Blocks) will constitute the shared ones. Therefore, there is an interplay between SDM-C and SDM-X since the SDM-C is pre-defining the amount of resources but based on the fact of sharing some of them (SDM-X role) we need to specify which resources will be shared. Therefore the algorithmic framework will have to take into account all pooled resources to find the optimal (and/or near optimal) for sharing so that a specific objective function is maximized/minimized.

3) **Tight Coupling:** Finally, the last case is when the resources for individual tenants as well as the shared ones are optimized jointly. Computationally this is the most complex one and in terms of architecture this case requires that the functionalities of the SDM-C and SDM-X overlap. In other words, if we would like to pool resources and share at the same time the roles of those two elements actually overlap since we need detailed information per tenant/slice. Note that in Case II the tenants provide a fixed set of shared resources and then SDM-X tries to optimize their use. On the other hand, in tight coupling these resources are not pre-defined and the operation of sharing takes into account all resources by trying to find the optimal subset for sharing. An interesting question that arises is how much we lose if we decompose this decision making between SDM-C and SDM-X. This is an issue that we aim to shed further light.

**VI. CONCLUSIONS**

The possibility of multiple services and tenants sharing the same network infrastructure is one of the main advantages proposed for use of SDN and NFV on 5G networks. However, virtualization and softwarization alone are not enough: resource sharing has to be an integral part of the design of any 5G architecture, and new ways to coordinate the usage of resources by many entities must be developed. In this paper, we described how the 5G NORMA tackled both of those challenges.

**REFERENCES**


