Securing network slices in 5th generation mobile networks

Luis Carlos Suárez Trujillo

DOCTORAT BRETAGNE LOIRE / MATHSTIC

Université de Bretagne Occidentale

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE BRETAGNE OCCIDENTALE

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Informatique*

Par

Luis Carlos SUÁREZ TRUJILLO

« Securing network slices in 5th generation mobile networks »

Thèse présentée et soutenue à Rennes , le 16 octobre 2020
Unité de recherche : Lab-STICC

Rapporteurs avant soutenance :

Maryline LAURENT Professeure, Télécom SudParis
Ahmed MEDDAHI Professeur, IMT Lille Douai

Composition du Jury :

Président :	Guy PUJOLLE	Professeur, Pierre et Marie Curie University (Paris 6)
Examinateurs :	Maryline LAURENT	Professeure, Télécom SudParis
	Ahmed MEDDAHI	Professeur, IMT Lille Douai
	David ESPES	Maître de conférences, Université de Bretagne Occidentale
Dir. de thèse :	Philippe LE PARC	Professeur, Université de Bretagne Occidentale
Co-Dir. de thèse :	Frédéric CUPPENS	Professeur, Polytechnique Montréal

Invité(s) :

Philippe BERTIN	Ingénieur de Recherche, Orange Labs
Cao-Thanh PHAN	Ingénieur de Recherche, IRT b<>com
Daniel MIGAULT	Senior Researcher in Security, Ericsson

Acknowledgments

I thank my family and close friends for all this long term moral support, counseling, love and good energy during my studies.

I thank my thesis directors and supervisors David Espes, Frédéric Cuppens, Cao-Thanh Phan, Philippe Bertin, and Philippe Le Parc, because of their patience, good spirit, precise advice, counseling and accurate direction during my doctorate studies. Besides learning in depth about a subject I thank them for teaching me how to learn.

I thank my supervisors, advisers and fellow colleagues at IRT b<>com, for their support, counseling, collaboration and good company during this epic and marvelous part of my life.

Table of Contents

List of Figures

List of Tables

INTRODUCTION

The telecommunication industry has shown a high level of dynamism from the last 20 years. Reports from key players in industry agree in the projection of two-thirds of the world population having Internet access by year 2023, each user with at least three devices connected to the IP network. The fixed broadband speed will get to 100 Mbps and 5G speeds will reach an average of 575 Mbps by year 2023 [1]. It is evident from these data that to support the increase of number of users connected to the network, the desired throughput, and its future state, it is necessary to enhance the network infrastructure. In consequence, it is necessary to grow the number of interconnections via submarine cables, boost the construction of data-center facilities and operator exchange points, and include new emerging and enabling technologies to that infrastructure. Services provided by telecommunication companies are no longer a luxury: they are seen at this moment as a basic public service, just as water or electricity, used for everyday tasks, from work to leisure.

As time advances, individual users, governments and industry rely on telecommunication networks to provide services to citizens, supervise processes, interconnect industries, commercialize goods, share feelings and even monitor the population. The Internet helps to add value to services and each day humans demand even more from this network, may be by asking for new functionalities or by asking for more capacity, performance and sense of security from it. Time is seen as a big constraint in a world where speed is mandatory, where a lost second is a business opportunity that is gone.

To cope with these demands, humans rely in all means possible to communicate and transmit information. We have developed and incorporated the technology to answer to those needs in a timely manner. For our work of this thesis, mobile networks have proved to be a convenient mean to transmit and receive voice and data in zones where there is network coverage from the mobile network operator.

In addition to the intrinsic characteristic of a communication network to deliver information, security aspects about how this information is handled and delivered by network operators and content providers is gaining traction and interest by the users, companies and communication service providers. The security considerations in the service delivery

chain are beginning to be studied from design phase and not as an add-on after the service is deployed or after security breaches are detected.

This change of paradigm has been motivated by the surge of security incidents in telecommunication companies and industries [2]. As an answer, a secure-by-design approach has been created, which seeks to minimize the security risks of a system by embedding tactical response mechanisms within the architecture of the service. Response mechanisms can be set up at network level or in the application level, permitting to enhance the security of the service. In consequence, this allows customers to trust the provider and its services.

As the capabilities, services and security of mobile networks advance by including more components and technology paradigms for their conception, the network inherits not only the qualities and new functionalities, but also the threats and risks of those constituent technologies.

From this point of view, next Section provides the motivation to get a deeper understanding about the challenge of securing the communications in mobile networks.

Motivations

Today, the mobile technology that is mainly used is 4G, which is addressed to be used by us, human beings in our smartphones. This technology suits very well the use case of mobile broadband, and is performing enough to send and receive email, watch videos at a decent resolution and with a tolerable delay to load the content.

Human imagination is unstoppable. As shown in Figure 1, we imagine new services, new business opportunities and new use case scenarios for which the current 4G model does not scale in order to guarantee the fulfillment of the requirements.

In order to get a better connected society and support more services, it is necessary to enhance the network capabilities, operation management, agility in the relation with the business and the interactions between Internet and the physical world. To make these new services a reality, the design of a new type of network is required.

This new evolutionary scheme is called 5G. As shown in Figure 2, it provides the capabilities to supply to customers and companies new services besides the mobile broadband use case. For instance, the new architecture provides the capacity to hold connections of thousands to millions of intelligent objects, a low latency to enable fast communication between machines, and enhanced bandwidth capacity to enhance connectivity and

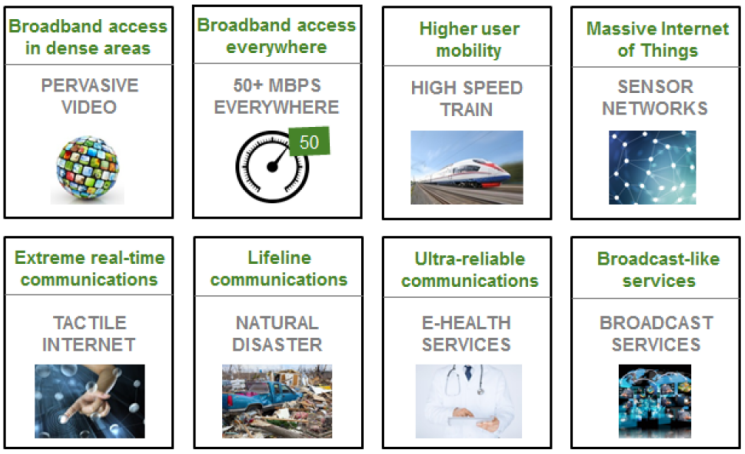

Figure 1 – 5G use case families and related examples [3].

multimedia experience.

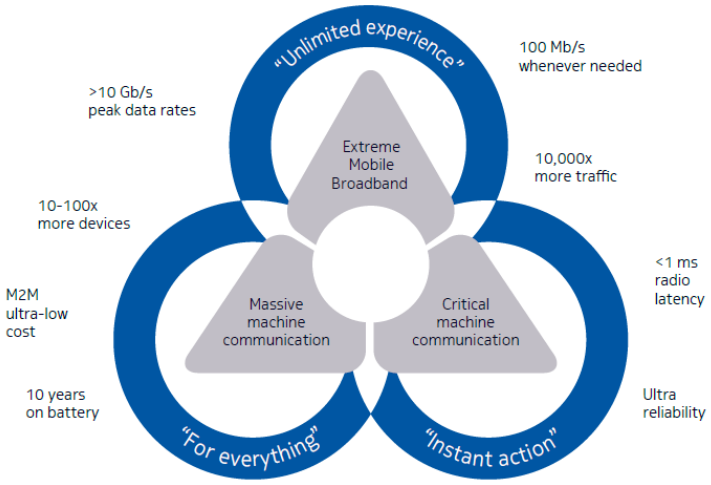

Figure 2 – Capabilities of 5G networks [4].

From a communication service provider point of view, the new architecture provides advantages, including:

— Agility to create and deploy new applications and services.

— Capacity to map the new interactions between the new players, new roles and new stakeholders.

— Opportunity to bring savings in Capital Expenditure (CapEx) – Operational Expenditure (OpEx) thanks to softwarization and virtualization technologies.

— Flexibility, thanks to its service-oriented conception, by the usage of programmabil-

ity and modularity, enabling it to change its configuration and morph into different forms according to the need.

Dealing with all these qualities has a consequence: it is very complex to manage. Specially considering that the use cases are dissimilar in configuration, different in the needed resources and, more important, it is impossible to have a single network that supports all the load and services. We could think about building dedicated networks to each one of the new services. But this approach is slow, prone to human error, with network segmentation schemes that do not scale and do not have the required speed to keep the pace with business and expansion of services. It relies a lot on human input and configuration, that hinders the speed that is required to deliver the service as needed. Besides these limitations, it would be very expensive in money and time to implement.

Another approach is to share the network and to segment it according to the needs. This is the foundational idea of Network Slicing. It is an enabler to deal with this complexity. It provides a novel segmentation scheme that addresses these requirements. But its realization poses challenges and problems to be solved, as is presented in the next Section.

Challenges for network slicing realization

As it will be presented in this document, Network Slicing is a concept that has been difficult to define. For now, we can say that a slice is a logical network serving a defined business purpose with specific characteristics, and comprises all the required network resources. Such resources can be physical or virtual, and either dedicated to a particular slice, or shared between several slices making use of isolation techniques. Its realization poses several challenges, for example:

— **How to orchestrate network slices?** This refers to how to deal with the complexity of the system, how to perform the management of agreement between stakeholders, how to map interactions between constituting entities, and how to provide mechanisms for Service Level Agreement (SLA) assurance.

— **How to implement resource sharing?** Resource sharing is difficult to implement because resources are heterogeneous and they are distributed in multiple sites, possibly under different administrative domains.

— **How to define isolation and how to enforce it?** The main idea is that resources are going to be shared between services according to their needs using virtualization

technologies. The challenge refers on how to let services use the assigned resources and deny the usage of resources assigned to other services. The difficulty translates into how to guarantee that resources are "separated" (depending of the layer in the service architecture) so that a compromised service does not affect another service that uses the same shared underlying infrastructure.

— **How to guarantee Quality of Service (QoS)?** QoS has to be specified end-to-end, with each equipment on the path from start to the end being capable of understanding the parameters that specify it and enforcing its compliance.

— **How to deploy secure network slices and their provided services?** The services that are delivered via network slices and the supporting underlying infrastructure must be secure. This topic is important because no operator would risk its infrastructure, image, services and in consequence, customers, by deploying services over insecure foundations. This challenge constitutes a big rough research question. This work is focused on the security aspects for network slicing.

Next subsection presents specifically these challenges related to security.

The research questions

Among the major security challenges for network slices, we find the following research questions:

1. *How to define and control the network slice behavior?* It refers to the parameters used to characterize the network slice in order to build its functional profile. Considered parameters can be, for example, its usual traffic patterns, destination networks for the traffic, mean processing load, and usual bandwidth utilization. With this information, guarantee that network slices behave properly among them.
2. *How to assure mutual authentication between actors?* Assurance that the entities that interact with a network slice are authentic and are truly the ones that are authorized to do so.
3. *How to control the access of users to network slices?* The Communication Service Provider (CSP) may provide several services using diverse network slices. The CSP must guarantee a connection from the user to only the required network slice. In the case that a User Equipment (UE) should connect to multiple network slices simultaneously, the CSP must ensure that the UE is not used as a bridge to route traffic between network slices.

4. *How to control multi level isolation?* Each constituting layer of a service can implement isolation, leveraging on the inherent qualities of that layer. For example, at a physical level, the service can use dedicated links; or at network level, a broadcast domain can be used to isolate traffic. The question relates on how to control and use those isolation strategies that can be provided at different levels of the architecture.
5. *How to secure Application Programming Interfaces (API)?* Open API are important for agility, openness, programmability. The CSP must ensure access to the API only from the authorized entities.
6. *What is the strategy to include accounting and non-repudiation?* Network slices can be instantiated in a multi operator, multi-vendor, multi-role environment. It is necessary to assign responsibilities, track actions and create interaction rules in order to know what is happening with the services.
7. *How to manage the security of the whole network slicing ecosystem?* Security-related data has "the 5V": volume, variety, value, velocity and veracity. Important tasks that need security data cover monitoring, fault detection and remediation, under fast changes of resource utilization.

It would be difficult to try to address all these challenges at once. Each one covers its own degree of complexity applied at different levels of the architecture, including the interactions that are needed between those layers and the dissimilar protocols and enabling technologies that contribute to its construction. It was necessary to choose a subset of the aforementioned challenges. As will be presented in Section 1.6, the emphasis of this research is on the inter-slice security and the management of the security for those network slices. This choice allows us to address the research questions number 2, 3, 4, and 7, which are relevant for a CSP that seeks to provide secured communication services that leverage on network slices for their implementation.

Organization of the document

The document is organized as follows:

In **Chapter 1** we will cover the state of the art of network slicing, its definitions, architectures and challenges. We will present our base architecture along with the isolated and secured by design concepts. This chapter led to two publications: **(i)** "Enhancing

network slice security via Artificial Intelligence: challenges and solutions" presented in Conférence C&ESAR 2018 [5]; and **(ii)** "Defining a Communication Service Management Function for 5G Network Slices", presented in EuCNC 2019 [6].

In **Chapter 2** we will present our first contribution that refers to the conception of an access control model for the 5G System, considered in the intra-slice use case. The access control requirements for the entities inside the 5G System will be presented, followed by the mathematical definition of the model. This work gave as a result two publications: one as a poster entitled "On an access control model enhancement for the 5G system" in EuCNC 2020 [7]; another as a full-paper entitled "Formalization of a security access control model for the 5G system" in NoF 2020 [8].

In **Chapter 3** the second contribution is described, consisting in the model for secure inter-slice communications. On it, we will show how enriched communication services are conceived thanks to inter-slice communication. We will present its mathematical foundation and how to perform the security policy validation. This chapter leads to a publication entitled "Managing Secure Inter-slice Communication in 5G Network Slice Chains" in DBSec 2020 [9].

Chapter 4 describes a novel method to classify the metrics for network slice isolation and measure the isolation between network slices. For this, we will present the proposed data structure for the metrics and the calculation process to quantify the isolation of one slice related to another. This work leads to the filing of patent number PCT/FR2020/050817.

Chapter 5 presents the conclusions of this work and future actions to develop.

Chapter 1

Network slicing

1.1 Introduction

CSP have been under a lot of pressure lately. All began with the diminution of the revenues for voice traffic, favoring voice over data traffic via Internet. They have been forced to create new sources of revenue by expanding the portfolio of services and customers. The number of mobile devices using these services has increased, creating a surge in traffic traversing the infrastructure. In addition, the new services have more demanding bandwidth and latency requirements, putting at risk the availability and stability of the network infrastructure. Capacity planning in order to trigger network expansion is painful to make, because it involves high costs to purchase new equipment to guarantee service to the customers. The reason is that the network is monolithic, dependent on specialized equipment and does not scale properly to address the variety of services that need to be provided [10] [11].

To solve these problems, several technological paradigms have emerged: Software Defined Networking (SDN), cloud computing and Network Functions Virtualisation (NFV). Individually, they have proved that they help to provide elasticity, resource sharing and optimization, programmability and automation to networks. But in order to fulfill the use cases that are proposed under the 5G umbrella, it is necessary to make them cooperate. The challenge is that each one addresses a different problem from the virtualization point of view (network abstraction, resource sharing and function abstraction, respectively) and a common ground for communication and global understanding of the problem is needed.

Network slicing is envisioned as a framework that, with the symbiotic relationship between these enabling technologies, helps to provide the promised services by composing the required resources and functions into a complete end-to-end service that meets the requirements of vertical business stakeholders. The challenge is to find the correct way to assemble these components and to have a proper communication among them, having in mind SLA, security management and corresponding Key Performance Indicator (KPI) to

monitor the end-to-end services.

Network slicing is going to be used in the different layers that compose a telecommunication network operator: **(i)** Access Network (AN), which covers the wireless or cable media to connect the final users; **(ii)** Core Network (CN), which contains the components that provide the control and functionality for the users; each one of them rely on **(iii)** an infrastructure layer, represented by resources; and **(iv)** a services layer, which finally offers a function or service.

There are several published papers that address the utilization of network slices at each layer. For example, regarding the *access* and *core* layers, authors in [12] provide an extensive survey stating the key principles for network slicing, indicating use cases, and denoting the mechanisms to enable resource sharing along with important challenges for its realization. Authors in [13] perform a revision of use cases and the state of the art in the *infrastructure* layer in relation to network functions, service, and the Management And Network Orchestration (MANO) component. These two papers also stress on the importance of enabling technologies such as SDN and NFV to build the network slicing concept for telecommunication networks. On the *services* layer, authors in [14] mention the use of network slicing as an enabler to multi-access edge computing and provide the functionality this technology requires. Finally, from a global point of view, authors in [15] provide a deep view on the end-to-end architecture for 5G, network slicing as its enabler, security and its management and orchestration.

The aforementioned literature explains the enabling technologies for 5G networks, network slicing among them. Despite of that effort, we detect that there is a lack of emphasis on the relation of the network slices with the new elements that would provide support information for the realization of this concept. At the same time, the justification of the usage of network slicing is not strong enough, as well as the challenges for its implementation. Encountered gaps consider the case that if a CSP chooses only one approach to implement network slicing, the CSP will lack of important features provided by other architectures. We want to make emphasis that network slicing is the key for CSP survival and realization of proposed 5G use cases: without network slicing the 5G vision will not become a reality and CSPs will struggle to keep up the pace to develop, innovate and deploy new services.

In order to fill the detected gaps, this chapter will be focused on network slicing (not 5G) addressing the issues from four angles. First, a state of the art about network slicing will be provided, with its definitions, objectives and properties according to Standards

Developing Organizations (SDO). Then, the architectures proposed by SDO will be presented in order to identify gaps and missing features, being their absence a major obstacle to deploy network slices. After this, major challenges will be described, covering different areas that are involved with network slices. Finally, a comprehensive definition and a generic architecture will be presented, which is compliant with the ones proposed by the SDO and enabling its implementation.

The importance of this chapter lies on the fact that, even though the proposed architecture does not solve all the problems, it supplies all the components and functionality to combine the qualities and requirements for network slicing from all SDO. This high compatibility makes the realization of the concept possible.

1.2 Network Slicing Definitions

1.2.1 Definitions provided by SDO

Coming up with a concrete definition of network slicing is a great challenge not only because of the abstraction level that it implies, but also because of the multiple definitions found in the literature. Some of them are provided by SDO which have their view of what a network slice is and what should be its purpose according to their business focus and the field of expertise e.g., mobile networks, Network Functions (NF) orchestration and resource management.

Next Generation Mobile Networks (NGMN)

The NGMN organization in the seminal 5G White Paper [16] describes the network slicing concept as an important part of the architecture of a 5G network. They provide a wish list about what a network slice should perform and what to achieve regarding deployment and business requirements. NGMN defines the network slice as an entity that wraps network functions and contributes to the realization of a communication service. For its realization, NGMN states that the network slice is derived from a template or blueprint that describes its internal parameters, configuration and operational specifications. They define its constituent layers (service instance, network slice instance and resource layers), provide a guideline about network slicing, interactions between providers, elaborate a use case about orchestration and management (focused on 5G) and provide as well a list of security issues focused on network slicing [17].

3rd Generation Partnership Project (3GPP)

3GPP is focused on the interaction between a UE and the functions that deliver packet services to that user, due to its expertise in mobile network field. 3GPP defines a network slice as a logical network that provides specific network capabilities and network characteristics [17], part of the high level architectural requirements for NGMN [18]. When deployed, it contains a set of network function instances and required resources in form of a network slice instance, which is built from a network slice template, representing the network functions and resources to provide telecommunication services. In addition, they say that a network slice is a new paradigm in which logical networks are created with the appropriate isolation, resources and optimized topology to serve a particular purpose or service thanks to enabling technologies such as NFV and SDN [19].

5G Infrastructure Public Private Partnership (5G-PPP)

5G-PPP (an organization that seeks to secure the European leadership in areas where Europe is strong or where there is potential for creating new markets) has a business driven focus. 5G-PPP defines a network slice as a composition of adequately configured network functions, network applications and underlying cloud infrastructure (may it be physical, virtual or emulated), all of them bundled together to meet the requirements of a specific use case [10]. Requirements, as specified by the customer, go in hand under the umbrella of a business purpose. The network slice is a way to carry out a service of a communication service provider portfolio, whose behavior is realized by a network slice instance in order to satisfy the demand from the customers. For 5G-PPP, network slicing is a framework with a multi-domain perspective, spanning management, orchestration, administration, security and technical fields with the objective of satisfying demand by **(i)** mapping the desired SLA to the resources; and **(ii)**, using automation in order to deal easily with the life cycle management of the network slice. Automation is necessary in order to deal with the end-to-end nature of network slicing, which covers all network segments.

Open Networking Foundation (ONF)

ONF, as an operator-led consortium that helps to realize the full potential of SDN paradigm, focuses on the management of the resources and providing abstractions via API, being these characteristics applied to 5G [20]. SDN is an enabler for network slic-

ing because **(i)** it permits a better usage of the infrastructure by the partitioning and assignment of a set of resources allocated to the network slice; and **(ii)** it provides a user customizable, application-aware view. These views permit a combination of all relevant network resources, network functions and network assets required to fulfill a specific business case, including Operation Support Systems (OSS) and Business Support Systems (BSS) processes. That is why for ONF, the establishment of a slice is business-driven and not technology-driven. A set of such dedicated resources can be called a network slice instance which must specifically support: **(i)** connectivity between endpoints; **(ii)** the resources to process traffic where it is required; and **(iii)** the network and operation management and business.

European Telecommunications Standards Institute (ETSI)

ETSI focuses on the Life Cycle Management (LCM) of network functions and their administration, due to its expertise in orchestration of resources. They do not provide a definition for network slicing, but they build the relationship between key components from other SDO and NFV in order to construct the network slicing concept [21]. Showing this relationship is important because network slicing will not only be a key component in 5G networks, but because the realization of the network slices will consist of NFV network service instances. In order to do so, the aforementioned document builds the mapping between network slicing concepts into NFV concepts and identifies potential gaps.

1.2.2 Objectives

According to SDO, network slicing seeks to satisfy the demand of the customers by performing: **(i)** SLA to resource mapping, so the CSP can guarantee performance for the services maintaining the economies of scale provided by resource sharing techniques [22]; and **(ii)** LCM leveraged with automation [23], to enable network elements and functions to be easily configured and reused to meet a specific end-to-end requirement (for example, a concrete latency or quality of service requirement on the core and access network). These two elements allow provisioning of network slices only with the needed resources [24]. This way, it provides a positive economic impact by the reduction of investment in unnecessary features and reduction of capital expenditure and operational expenditure. As a consequence, the CSP increases network revenue [20]. Network slicing will permit

the growth of the capacity of the mobile network and offer new services, going beyond the current static approach. It will also integrate the Operations, Administration and Management (OAM) tasks as part of a 5G network [3].

Since a network that supports simultaneously all use cases and its performance requirements is difficult to design and maintain [23], network slicing is a key design strategy in order to improve: **(i)** network capabilities (data rate per user, end-to-end mobility); **(ii)** operational sustainability (automation to enhance self-organizing network approach); and **(iii)** business agility (to support more services besides mobile broadband) [3]. These improvements help to realize the use cases envisioned for 5G, which are grouped into basic service classes [10]: enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC).

1.2.3 Characteristics

In order to achieve its objectives, a network slice has several characteristics that identify it. The characteristics vary according to the point of view of the SDO.

NGMN does not provide a key characteristic for a network slice. They state the guiding principle by which they leverage on the sharing of resources to fulfill its objective.

For 3GPP, network slice characteristics are around two topics. The first one refers to the management and orchestration [19] which are centered on the network slice LCM and covers fault, performance, configuration, policy, data isolation, and multi-operator and automation management. The second one refers to service characteristics [17] which focus on the description of the parameters of operation, such as functionality, performance, isolation and how they interconnect with the rest of the components of the Evolved Packet System (EPS).

The characteristics of a network slice according to 5G-PPP [10] are related to **(i)** the tight interaction between SDN and NFV; and **(ii)** the integrated Fault, Configuration, Accounting, Performance and Security (FCAPS) management in single or multiple administrative domains in an end-to-end fashion. The characteristics rely on resource sharing techniques such as, multitasking, virtualization, and multiplexing, permitting to **(i)** decouple the functionality from the resources needed to execute the function; and **(ii)** partition the resources into isolated execution environments. These characteristics enable the partition of all components in the network, resulting in end-to-end slicing.

In the ONF point of view, the characteristics of network slices are around resources [20], which **(i)** must be divided and assigned to be used in isolated / disjunctive / shared

manner; and **(ii)** must be properly defined so they can be abstracted via the northbound interface to OSS/BSS, to guarantee technology independence. The abstracted resources are exposed via APIs, which have to be well defined, (so services become programmable) and must be designed as modular, re-usable building blocks with model driven and callable interfaces. Something interesting about this approach is that the two perspectives SDN has about the architecture (resource oriented or service oriented) make it the center of a feedback loop between the OSS/BSS and the resources, leveraging on client/server interaction.

For ETSI, the key principle regarding network slicing is the isolation [21]. In order to achieve this, ETSI leverages on SDN capabilities to propose one architectural example that uses two SDN controllers: **(i)** infrastructure SDN controller; and **(ii)** tenant SDN controller, realizing this way the desired isolation. This proposal can be seen in [21].

Table 1.1 shows a summary of the network slice characteristics presented in this section.

Table 1.1 – Summary of the characteristics for network slicing according to each SDO.

SDO	Characteristic	Description
NGMN	Generic vision	Provides guiding principle for its purpose.
3GPP	Management and orchestration	Centered on network slice LCM.
	Service	Describes parameters of operation
ONF	Resources	Divided and assigned. Abstracted properly and exposed via APIs.
5G-PPP	SDN and NFV interaction	Rely on resource sharing techniques.
	Integrated FCAPS	Rely on resource sharing techniques.
ETSI	Isolation	Leverages on SDN capabilities to achieve this characteristic.

1.3 Architectures proposed by SDO

Standards have been viewed as a mechanism to avoid vendor lock-in [25]. Specifically, standards provide **(i)** a broad agreement over a well-defined scope; **(ii)** a well-accepted

policy and intellectual properties guidelines; and **(iii)** commands authority over the topic being standardized [26]. These are sufficient reasons to consider SDO as important players leading the way of how a technology should be used and influencing other actors to follow those guidelines.

The purpose of this section is to provide a vision of the network slicing architectures proposed by the previously mentioned SDO and make a useful comparison among them.

1.3.1 NGMN

As presented in Section 1.2.1, NGMN is the SDO that owns the network slicing vision. Their generic definition and use cases reflect the intention and what is desirable when using a network slice. NGMN proposes a three layer architecture to realize the network slicing concept. It is illustrated via an example shown in Figure 1.1 which points out a typical near-future scenario for a CSP that provides five services to customers. Each one of them leverages on one or several network slices, customized according to the service requirements. For instance, the architecture depicts the service deployment for a simple

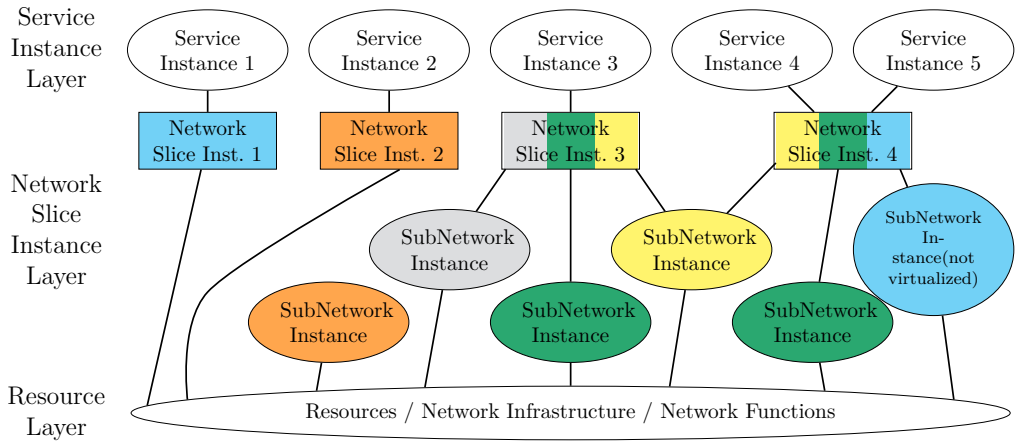

Figure 1.1 – Network slicing conceptual outline, according to NGMN [27].

case like Service Instance 1, which uses the Network Slice Instance 1. Increasing complexity, Service Instance 3 leverages on three Sub-Network Instances that make up the Network slice instance 3. A more complex case is depicted for Service Instance 4 and 5, which both share the Network Slice Instance 4. NGMN even covers the case in which the yellow Sub-Network Instance is shared between Network Slice Instance 3 and 4. This separation in diverse network slices (or even network slice subnets), allows the CSP to

have manageability of the service, to make them run as desired, to provide optimization of the service and manage their interactions and security.

The layers that comprise the architecture are: **(i)** service instance layer, which represents the services, denoted by a service instance; **(ii)** network slice instance layer, which provides network characteristics which are required by a service instance; **(iii)** resource layer, which covers resources (physical or logical) on the network infrastructure. For NGMN, a service instance refers to the realization of an end-user service, made possible thanks to a network slice. This service instance is composed of network slice instances, which group together network functions and required resources to meet a characteristic of the service instance. NGMN specifies further that the network slice instance layer is created via network slice blueprints. Blueprints are created during design (or configuration time) and contains the description of the structure, configuration and the instructions to perform the LCM of the network slice instance [27].

1.3.2 3GPP

According to 3GPP, network slicing enables a CSP to create customized networks which are optimized to provide solutions for different market scenarios. These scenarios usually demand diverse requirements to fulfill the service along different axis, such as the areas of functionality, performance and isolation. 3GPP adheres to the three layer scheme depicted by NGMN for network slicing (service instance, network slice instance and resource layer) and adds management functions for this architecture [28].

3GPP in [18] proposes an architecture to enable next generation systems to support network slicing, which is represented via an example shown in Figure 1.2. On it, 3GPP shows three network slices (A, B and C) which contain NF that can be specific for each network slice or shared as common NF. The purpose of this example is to demonstrate the functional entities that are needed to guide a UE towards a desired network slice. The basic components of the proposed architecture are:

- Subscriber repository function: contains subscriber subscription information, UE usage type, service type and UE capabilities.
- NF, which can be:
  - Slice selection function (SSF): selects an appropriate slice for the UE based on the UE's subscription information.
  - Common control plane NF (CCPNF): CP entry function, which at least includes the MM (Mobility Management) function, AU (authentication) function,

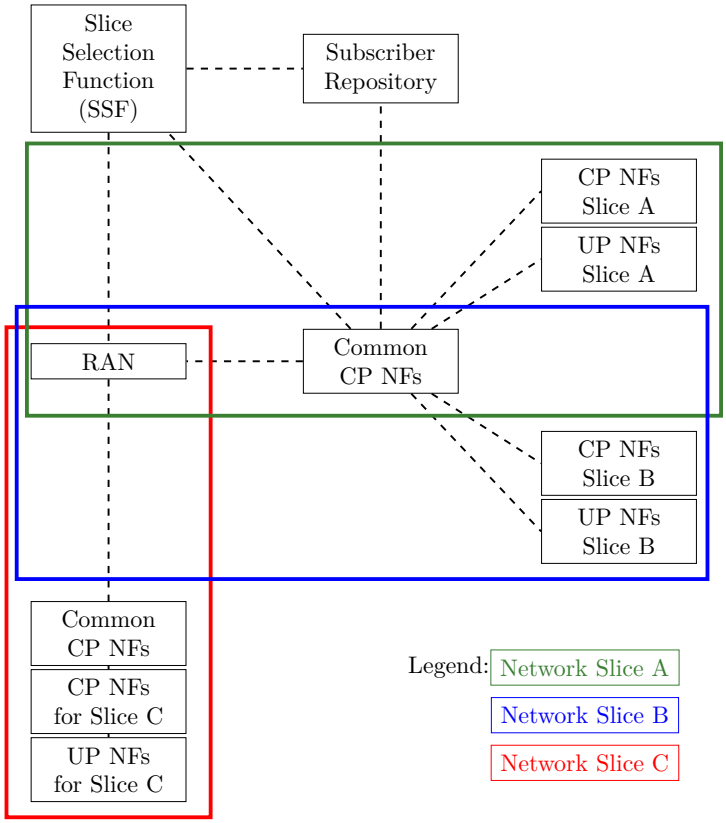

Figure 1.2 – Control plane architecture for network slicing according to 3GPP [18]. Network Functions (NF) are distributed on the Control Plane (CP), User Plane (UP) and Radio Access Network (RAN).

and Non-Access Stratum (NAS) Proxy function. The CCPNF is shared among different slices.

— Slice specific CP-NF: the NF which are located on the non-shared slice parts.

3GPP specifies the interactions between the UE and the core network (network slice selection, network slice instance selection, re-selection and association, among others) and how would these elements be managed. The steps to perform the LCM of a network slice instance are enumerated covering from the preparation (design, pre-provision, network preparation) to its instantiation, configuration and activation. A supervision scheme is also considered along with the decommissioning phase, where the network slice instance is deleted if it is no longer required [19].

3GPP divides the qualities of a network slice into two working domains: management and orchestration [19]; and service domain [17].

Management and orchestration starts when all the steps to perform the LCM of the

network slice are finished and in consequence, the network slice is instantiated. Since the details of these steps are outside the scope of 3GPP, external entities would execute the specified process into the infrastructure. It covers the network slice fault, performance and policy management along with the scheme to isolate data and automation. The entities shown in Figure 1.3 are in charge of executing the aforementioned management and orchestrations tasks. Each one of them provides a management function as follows:

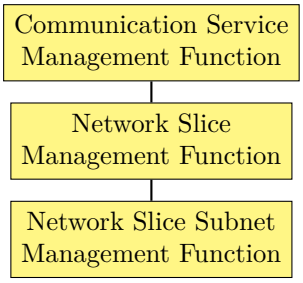

Figure 1.3 – Network slice related management functions, according to 3GPP [19].

— Communication Service Management Function (CSMF): Acts as a translator from the communication service related requirements to network slice related requirements.
— Network Slice Management Function (NSMF): Performs management and orchestration of the network slice instance and derives network slice subnet requirements from network slice requirements.
— Network Slice Subnet Management Function (NSSMF): Responsible for management and orchestration of network slice subnet instance.

About the *service domain* that is provided via the network slice instance, 3GPP suggests several criteria to customize and optimize a network slice with parameters such as functionality, performance and isolation at several levels and types (security isolation, resource isolation, OAM isolation). Other important tasks focus on mechanisms to perform identification and selection of a network slice, roaming support and how to internetwork with the existing EPS. These topics are important for 3GPP because their interest is not only to provide interoperability with existing networks and other operators, but to specify a detailed procedure for the UE to supply the core network with valuable information to select the most appropriate network slice for the required service.

1.3.3 5G-PPP

For 5G-PPP, a network slice helps to realize a service requirement of a service provider portfolio. They leverage on network slicing to utilize resource sharing technologies and "softwarization" techniques to deliver a fully decoupled end-to-end network. 5G-PPP does not provide a stand-alone architecture for network slicing, better, they make it part of their proposed 5G architecture, which is shown in Figure 1.4. The figure shows the original

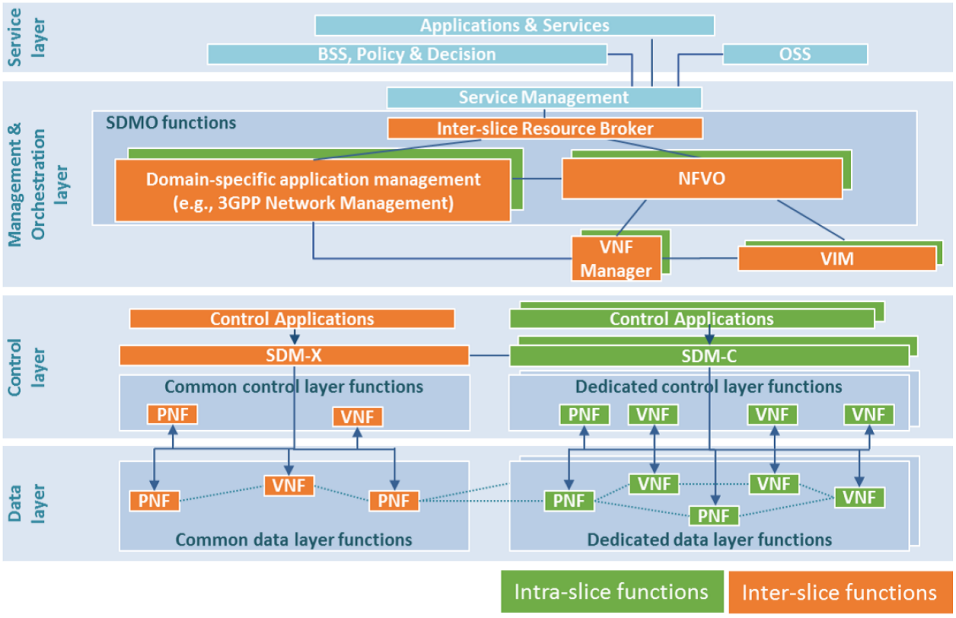

Figure 1.4 – Architecture functional layers for 5G, considering network slicing, according to 5G-PPP [10].

architecture which is at first sight difficult to understand due to its multiple components. What is important to keep in mind is that it is divided into layers for **(i)** the service (with business-purpose and ruled by policy); **(ii)** the management and orchestration of the network slices (leveraging on 3GPP and ETSI components for this purpose, including the inter- and intra-slice approaches); **(iii)** the control of applications (over dedicated and shared network functions powered by SDN); and **(iv)** the data layer (where Virtual Network Functions (VNF) and Physical Network Functions (PNF) process user data traffic).

Specifically, among the layers in which this architecture is divided, network slicing takes part in the management and orchestration layer within an entity called inter-slice resource broker. This entity would receive resource-facing service descriptions from the

service management entity, which acts as a translator of the customer-facing service descriptions received from the service layer. Conceptually, 5G-PPP establishes two types of network slicing services, according to the desired level of control provided to the customer. The first one has to do with the provisioning of *virtual infrastructures (VI)*, which requires direct hardware support and follows the Infrastructure as a Service (IaaS) model for the creation of network slice instances. The second type refers to the provisioning of *tenant's owned network services* instantiated over a shared infrastructure. These network services refer to VNF connected among them via VNF Forwarding Graph (VNFFG). The whole service would be specified thanks to VNFFG - VNF - Network Services (NS) descriptors and, through the exposure of API, the degree of impact in the control and management of the network slice for the customer can be specified.

1.3.4 ONF

The SDN architecture conceived by ONF (and used extensively by network providers) is an enabler to support multiple client instances over a common, shared infrastructure. Resource virtualization and recursion, as key SDN concepts, grant the desired flexibility that network slicing concept seeks to provide. ONF does not present a network slicing definition by its own means, but they do furnish the tools to power the following two views for network slicing: **(i)** the business view, in which all the required resources to fulfill a business case are combined accordingly i.e., from a bottom-up approach, SDN provides "infrastructure universality"; and **(ii)** the technical infrastructure view, in which SDN allows to partition and assign resources that can be used in isolation or shared. From a top-down approach, it would mean that SDN enables a user customizable and application-aware view of the resources.

Figure 1.5 shows an extract of this architecture to illustrate ONF's view. The SDN architecture relies on client-server relationships to link the resources, controller and applications, along with an SDN controller at the center of a feedback loop to act as a mediator of client's requirements. The SDN controller works with two types of resource views: **(i)** the client context, used to interact with the client, with all attributes of a service as requested by the client; and **(ii)** the server context, which contains everything necessary and sufficient to interact with underlying resources.

Adapting these concepts to out network slicing topic, the SDN client context would have the same behavior as a network slice because it provides the complete abstract set of resources and supporting control logic for constituting a slice, including the complete

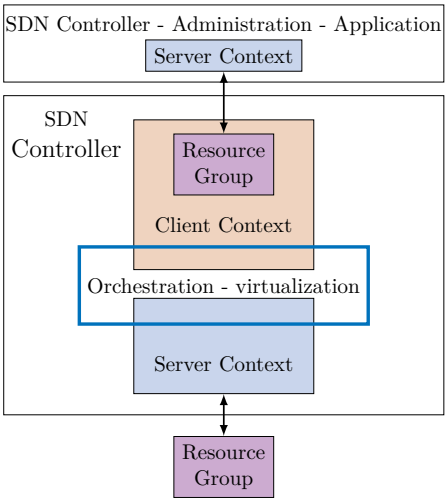

Figure 1.5 – SDN-based Network slice abstraction, adapted from [20].

collection of related client service attributes. Inside the client context, there is a resource group entity, which defines the semantics of the interfaces presented to the client.

The top layer constitutes any application whose purpose is to manage the resources via the exposed services of the client context. This application can be related to administration, to a specific service or even another SDN controller.

According to ONF, the client context maps to the concept of NFV-NS. The control of the SDN architecture is complemented by the ability to support network slice blueprints, which contain predefined information (in terms of services and abstract resources) to fully define the network service [20].

1.3.5 ETSI

ETSI does not propose an architecture of their own, but as an enabler to perform LCM of virtual resources, it is used under several frameworks.

As presented previously in Section 1.3.3, 5G-PPP uses ETSI NFV MANO architecture to enhance the management and orchestration system used for 5G architecture. It is also used as a type of network slice service which is instantiated directly over the shared infrastructure, leveraging on ETSI's definition of network service.

For 3GPP, 5G requirements specified in [29] (like network slice LCM, its elasticity and adaptation of capacity and resiliency) are supported using ETSI NFV concepts. For instance, ETSI merges the network slice management functions of 3GPP (in yellow) with NFV-MANO (in green) and its resources (in purple) [21], as shown in Figure 1.6.

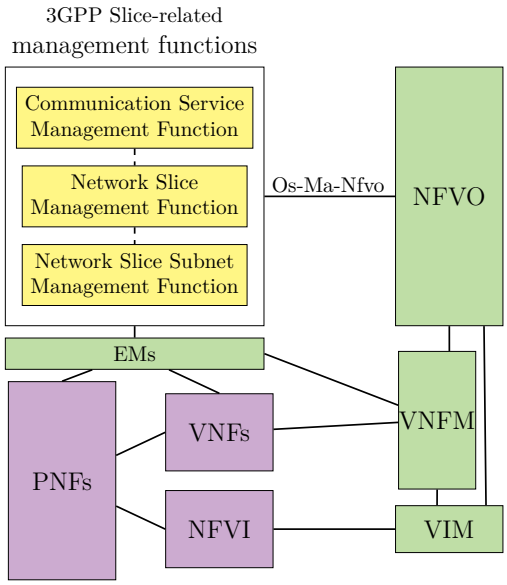

Figure 1.6 – Network slice management in 3GPP within an NFV framework.

The requirement for this configuration is that there must be a translation of the template parameters from 3GPP model to descriptors on ETSI model. It is also necessary that 3GPP slice-related management functions comply with the message format in order to communicate correctly with ETSI NFV Orchestrator (NFVO) via the Os-Ma-Nfvo interface.

This architecture is different to the previous ones that have been explored, because there is no view of a service layer. It only covers a section to manage the network slices; the management of the network services with its VNF and PNF; and the infrastructure over which the functions are instantiated.

1.3.6 Comparison

Since each SDO proposes different architectures to construct the network slicing concept, it is useful to compare them to identify common building blocks and assess trade-offs between them. This is summarized in Table 1.2. As interesting facts of this comparison, 5G-PPP provides a good starting point by showing the need of an entity called inter-slice resource broker that manages the cross-slice resource allocation for network slices. This broker allows to have a broader visibility than the one provided by 3GPP's management functions. MANO's approach proposed by ETSI is necessary because a LCM scheme is required for the network services, which are the components of network slice instances.

Table 1.2 – Summary of characteristics and missing qualities related to network slicing.

SDO	Characteristics	Missing qualities related to network slicing
NGMN	• Has the vision for network slicing. • Provides user stories to illustrate its usage and support the provided definition.	• Since the description of the vision is provided, there is no insight about building blocks or major functions.
3GPP	• Very elaborate, as they know how to develop the UE to CN interaction in LTE. • Provides detail about the division of tasks in Control Plane: common and custom functions.	• It requires new parameters to specify requirements: This involves adding new data structures, and the use of more signaling.
5G-PPP	• Uses explicitly components from NFV architecture (NFVO, VNFM and VIM). • Includes common and dedicated control layer functions, like the ones that specifies 3GPP.	• The architecture does not explicitly say how the layers interact. • VIM sends commands to the infrastructure via the control applications: this could increase signaling traffic and delay in deployment.
ONF	• Owners of the SDN vision. • Provides an extensive view of control plane functionalities to enable network slicing. • A VNF is a resource for a SDN controller. • SDN controller behaves as an intermediary controller unit between OSS/BSS and NFV MANO entity.	• Still no standard way to express the requirements from applications to controller (NBI protocols). • No specification on how to segment the resources from infrastructure providers. • It lacks of the competences to manage the life cycle of network slices and its internal resources.
ETSI	• Owners of NFV point of view (Applications, MANO framework, and infrastructure). • Provides lifecycle control of VNF. • Easy interaction with SDN paradigm. • Interaction with 3GPP view about management functions.	• No slice management functions. • No clear isolation scheme. • No support for multi-site deployment. • No scheme for resource allocation prioritization.

Regarding ONF, the proposed controller with an administrative context and server context to manage applications and resources provided to clients is a great scheme to provide addressing, traffic and namespace isolation, not only regarding the exposure of the management functions, but also the representation of the resources for the customer and the visualization of those resources for the owner.

Another way to compare the SDO is in relation to their impact. For example, NGMN would be classified as a **visionary**, because they provide the idea and the model for

network slicing. Another category considers the **standard enacting organizations** such as 3GPP, ONF and ETSI. The slight difference among them relates that ONF relies on existing interfaces (their model adapts to any architecture since they see other entities as clients and servers), while 3GPP and ETSI considers the creation of new entities and interactions between them. This approach would also cover 5G-PPP, not only because they hold a vision but also acts as an integrator in order to push the development of the technology. Their proposed architecture portrays more complex functionalities due to the fact that seek to integrate a large amount of use cases.

From the ideas exposed in this section, it is concluded that none of the existing approaches are sufficient in order to fully build a network slicing definition and architecture. Even more, it is necessary to point out the challenges that are brought by this new concept. This will addressed in the next section.

1.4 Challenges in network slicing architecture

This section addresses the challenges that need to be overcome in order to build a network slice, enhance their security and provide better ways to manage them. The challenges have been grouped according to categories to make them more understandable. Even though it is utopian to think that a single architecture can solve all problems, elements for an enhanced architecture are presented as a complete ecosystem, in order to try to solve the issues shown in the following subsections.

1.4.1 Orchestration and management

Orchestration and management constitutes a challenge as networks and services grow in size and complexity, not only of the network functions but of the interactions between components and stakeholders. Heterogeneity of the infrastructure is commonplace, so we need tools to identify domain boundaries belonging to different administrative domains, providing this way a complete multi-domain orchestration to serve the desired functionalities [30]. This challenge gets more difficult if the service sends requirements for resources very frequently (in short timescales), which poses a challenge to the speed of the policy validation while permitting cooperation between management functional blocks and the resource orchestrator [31]. Since physical resources are finite, cooperation is needed between public and private clouds in order to scale up the resource pools as needed, usually

at peak times. For this, it is necessary to enhance trust mechanisms, provide cross domain knowledge of the resources, and revise security and administration policies [32].

Enhancing the orchestration capabilities would permit to manage better end-to-end services, nonetheless there is a blockage related to the lack of a common language that standardizes the service definition [13]. Lawful interception is also a challenge, that could be tackled by having higher granularity of NF at expense of the effort to chain those resources [33].

From the point of view of the application, the challenge is to design a mobility management scheme that is aware of the application running on a network slice, so we can customize the response according to the needs and speed of the required handover, in order to preserve the QoS for the user [34]. To make sure that requirements are fulfilled, standardized API and protocols are required to provide seamlessly network slice and service performance monitoring [33] [30].

1.4.2 Resources

Challenges about the resources are important because of their heterogeneity, dissimilar location for hosting them, different administrative domains, and the different entities that could offer them as a service to their customers. Due to the fact that most of resources will be virtualized and shared between customers, challenges cover how to guarantee performance of the services over shared infrastructure [31], the dynamic behavior among slice tenants [34] and its fairness [35] in order to make network resource usage more efficient. Dynamic control of the resources is required in order to provide a stable environment to the instantiated network functions [36]. Since VNF rely on the physical infrastructure, it is desirable to know where they are instantiated physically in order to have control of the propagation of physical layer failures to the virtual resources [37].

1.4.3 Isolation

Isolation is a complex topic because it can be applied to different layers of a network architecture in order to provide secure capabilities to the network slice.

One important challenge refers to *inter-slice isolation* to assure control plane and user plane complete independence. This would help to **(i)** guarantee that an attack on a slice will not affect other network slices, especially in cases when certain control functions are shared [34]; **(ii)** guarantee that, when a UE is connected to several network slices at

the same time, it does not filter traffic from one slice to another [38]; and **(iii)** avoid the propagation of an attack between network slices when network functions are shared, known as cascade effect [37]. Access to those common resources must be managed intelligently as suggested in [22]. These isolation requirements must be formalized and standardized to permit a proper definition of its parameters along with the proper values to achieve their objective [33]. It is worth to mention that in [39] authors provide a first approach to solve this challenge, but since the parametrization covers multiple aspects and details, it is necessary to continue elaborating on this issue.

Another challenge is related to *intra-slice isolation*, which refers to the behavior of the components inside the network slice. As suggested in [34], intra-slice management could be implemented by running a virtual manager function as part of the slice, but it is necessary to better evaluate the parameters that would govern the isolation management. One alternative could be to consider the number of VNF per physical server that can be instantiated for a network slice [40].

1.4.4 Security

Security has been a trending topic involving applications and the network over which those applications are supported. It spans all layers of a network architecture including the user. One important challenge refers to the creation of a scheme to profile the behavior of a network slice. This is difficult because of the plethora of use cases and their characteristics, but this profiling would help to guarantee that slices behave properly and to avoid certain attacks such as Distributed Denial of Service (DDoS) or inter-slice traffic without authorization [38].

Due to the multiple actors and stakeholders involved in 5G, it is required to have assurance that the elements that interact with a network slice are truly the ones that are authorized to do so. This poses a challenge on the scalability, performance and maintenance of mutual authentication mechanisms, in order to avoid impersonation attacks against a network slice instance and different network slice managers [38] [41]. Following on this multi-operator, multi-vendor environment for 5G, it is necessary to map all the interaction rules between these entities in order to provide proper accounting and non-repudiation of actions and decisions regarding a service [33].

Since a CSP may provide several services via different network slices, the challenge is to allow a UE access only to the intended slice and deny access to other slices. One approach could be to instruct the RAN to allow the UE to connect only to the required

CN slice [33]. These services are enabled thanks to API, which ensure agility, openness and provide network programmability. The challenge is to guarantee security and privacy of the API without sacrificing usability [31].

Regarding the resources for the services, it is expected that they have different security levels and policies since they could belong to different providers. One challenge is how to enforce security of the network slice in order to avoid impact when a shared NF or shared slice is compromised under shared infrastructure [34].

1.4.5 QoS

QoS is a central premise into the value proposal for a communication service. Regarding network slicing, besides adapting constraints to consecutive paths when chaining a service via different CSP, the challenge is to take into account the dynamic nature of the infrastructure and network functions according to the requirements of the use case [35]. The difficulty relates on how the changing conditions in infrastructure involve a recalculation of the QoS and negotiation with other administrative domains in order to assure the agreed level of service.

Another challenge involves the creation of schemes that leverage on the advanced parameters used to specify QoS and provide better way to have slice differentiation according to the use case, such as eMBB, URLLC and mMTC [42]. Exploring this concept further, network slicing should not be seen as an advanced QoS mechanism. Network slicing leverages on the customization of functions and procedures to enrich the experience and service, such as control plane procedures for mobility management, location tracking, session establishment, among others in order to tackle a specific use case scenario. This customization level goes further than solely the QoS concept.

1.4.6 Radio access

Network slicing has an end-to-end nature, meaning it covers from the access network to the equipment where service is delivered. Radio access is a key component in an end-to-end network slice, since it is usually the first medium that a UE uses to access services from a CSP. One challenge relates to radio access collaboration, that is, how to aggregate multiple radio technologies in a cooperative way with the aim to deliver seamless mobility and higher throughput [43]. One difficult topic in this area would be the handover between technologies and the election of the best medium to achieve a desired connectivity quality

according to the use case, for example, choosing between 3G or 4G radio access while a vehicle moves through different radio-coverage zones.

Another important challenge is about dynamic spectrum sharing in order to optimize radio resources via virtualization and keep up with the dynamicity of the rest of the network. This would lead to a Radio as a Service (RaaS) approach [13]. Since spectrum is an expensive and limited physical resource, its sharing techniques are limited and must be optimized.

1.4.7 Other challenges

In the literature, authors call attention to provide backward compatibility with legacy technologies with the intention to guarantee gradual migration to 5G networks. This requires an interaction between regulation entities and manufacturers to ensure a profitable and easy transition towards 5G [31]. Since mobile operators have executed huge investments in 4G-LTE deployments, it is desirable to have a system that is fully compatible with it and provides further revenues from it.

Another important challenge relates to the level of granularity that should be used when decomposing a network service. The usage of fine-grained network functions poses an interesting challenge because of the trade-off between how easy it is to compose of a service with respect to the number of network functions to connect together. As expected, complete and complex functions are easier to manage and connect with each other but flexibility of service composition is lost [44].

1.4.8 Final remarks

Besides the dissimilar definitions, there are important challenges that must be addressed in order to create the network slice concept. Table 1.3 outlines the described security threats to network slicing, which are the compromised security objectives and which are the potential functional groups or entities that can help to solve the issue. Keeping in mind the impact of these challenges is important, because it helps to know the current limitations of implementations and is useful to solve them gradually as solutions gain complexity. As the major contribution of this chapter, the next Section addresses these shortfalls by proposing a full definition and generic architecture.

Table 1.3 – Summary of threats to network slicing, compromised security objectives and envisioned entities that can help to solve the issue.

Representative threat	**Security objective**	**Derived problem**	**Addressed by**
Elevation of privileges.	Authorization.	Execution of non-authorized procedures.	MANO, OSS.
Spoofing of identity: against physical platform, the slice manager and orchestrator.	Authentication, integrity.	Impersonation of infrastructure, managers, and network slice instances.	NFVO, VNFM, VIM, SDN controller.
Tampering.	Data Integrity.	Modification of data in transit and at rest.	VNFM, VIM, VNF.
Slice jumping and side channel attacks.	Privacy, authentication, access control.	Traffic leaking from one slice to another, via a user equipment. Extract data by observing network activities.	NSSF, CSMF, NSMF, MANO.
Insecure inter-slice communication.	Confidentiality, access control.	Information Disclosure.	CSMF, NSMF, NFVO.
Resource sharing.	Access Control, authentication, data privacy.	Lack of trust in the usage of the shared resource.	MANO, VNF, SDN controller, resources.
Usage of rogue infrastructure.	Trust.	Usage of resources that are not approved by the service owner.	MANO and SDN controller using Trusted Computing Base (TCB) capabilities.
Repudiation of actions.	Non-repudiation.	Deny the consumption of services or requests to change configuration of services.	BSS NBI towards the customer, SDN controller, MANO.
Denial of Service.	Availability, authentication, monitoring.	Unavailability of services and mangement capabilities. Bypass of security controls. Unusual traffic behavior.	OSS, MANO.
Resource exhaustion.	Availability, authentication, monitoring.	Network slice can access more resources than the ones permitted.	CSMF, NSMF, NFVO.
Lack of isolation of network slices.	Confidentiality, data integrity.	Problems in performance and security of network slices due to usage of shared infrastructure.	CSMF, NSMF, NFVO, SDN controller, resources.

1.5 Comprehensive definition and proposed architecture

In order to solve the architectural gaps, this chapter presents a comprehensive network slicing definition and a proposed architecture for its realization.

1.5.1 A comprehensive definition

The importance of providing a comprehensive definition is to merge the different views from the diverse proponents and reach a common understanding of what a network slice should cover and comply with.

Definition 1. A network slice can be defined as: "a logical recursive entity, which via enabling technologies and orchestration mechanisms, provides the necessary means to realize a complete service, that meets the requirements set by a concrete use case scenario powered by a business purpose and optimizing the resource usage".

Specifically, *recursiveness* permits a communication service customer, who consumes network slices from its CSP, to become a network slice provider, under the condition that the proper management tools are granted by its CSP in order to manage the network slice.

Enabling technologies help to manage the resources that compose the network slice. *Orchestration mechanisms* provide the means to **(i)** arrange the necessary NF inside a network slice; **(ii)** connect together the network slices in order to build the required network service; and **(iii)** monitor their behavior and keep track of their life-cycle management. All these responsibilities rely on the use of blueprints or templates that provide programmability and agility to the implementation of a service by specifying its operational parameters. The parameters of a network slice can be **(i)** specified by the customer at instantiation time, or **(ii)** defined according to a pre-configured service based on the use case.

Finally, *business purpose* is empowered by the agreement of KPI between the customer and the CSP, whose enforcement motivates the CSP to perform the necessary tasks to ensure the service is delivered as agreed.

1.5.2 Generic architecture

As shown in the previous sections, a single approach to network slicing is not sufficient, so it is necessary to select key components to extend and enhance the architecture. A unified view is provided, ensuring compatibility and interoperability with the functions and specific qualities proposed by each SDO. Considered components are: **(i)** a Business Support System (BSS); **(ii)** an Operation Support System (OSS) block that contains Communication Service and Network Slice (Subnet) management functions (CSMF and

NS(S)MF); **(iii)** an Operations, Administration and Management (OAM) block containing Fault, Configuration, Accounting, Performance and Security (FCAPS), Policy, Management And Orchestration (MANO) capabilities and controllers; and **(iv)** the resources. Their functions are explained in the following subsections and their disposition is shown in Figure 1.7.

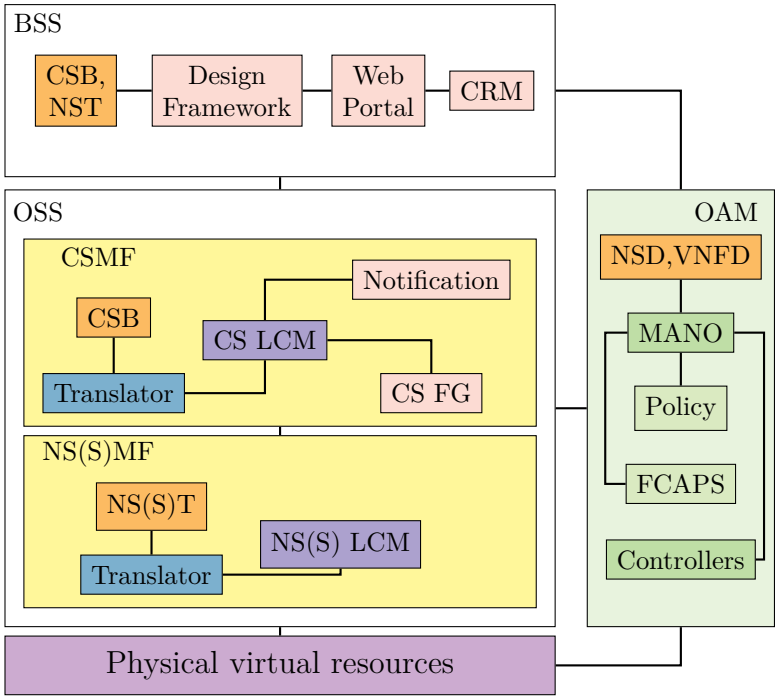

Figure 1.7 – Proposed architecture to solve the network slicing challenges.

BSS

This business support layer contains a *Design framework*, which allows the customer to fully specify the service. It provides the tools to specify their own VNF and modify the NST according to their service needs. Tools used by this framework that will accomplish this objective cover:

— Communication Service Blueprint (CSB): It contains pre-designed communications services comprised on the CSP portfolio, previously tested and ready to be used by the customer.

— Network Slice Template (NST): consists of a list of parameters needed to fully specify a network slice. In conjunction with the information inserted by the customer, it constitutes the network slice descriptor used during the instantiation process of

the network slice. The network slice template leverages on concepts from ETSI NFV such as VNF descriptor, VNFFG descriptor and network service descriptor, which are detailed in [45].

Functions from the Design framework could be performed by existing platforms such as the Open Network Automation Platform (ONAP) in its design-time section [1].

In addition to the *Design framework*, a *Web portal* is used as a tool through which the user can perform network slice and network service design. It contains available network functions and pre-designed services to ease the network slice deployment. Moreover, a *Customer Relationship Management* (CRM) software implements functions to manage customers, orders, products and revenue.

OSS

The operations block is inspired on the 3GPP approach. The *CSMF* matches customer requirements coming from the BSS to network slicing resources. It plays a key role to manage the life-cycle of the Communication Service (CS) (performed by the CS LCM block in Figure 1.7) and it is capable to have an end-to-end vision of the service. To do so, it uses a Communication Service Blueprint (CSB) which is used by a Translator block as a mapping strategy to find the suitable network slices to provide a communication service. Beyond finding the appropriate network slices, the CS LCM block will link them using the Communication Service Forwarding Graph (CS FG) block, in order to instruct lower layers how to connect the slices and create the complete service. A Notification block is used to send information to the BSS, so the customer and CRM are informed about the situation of the offered service.

The *NSMF* and *NSSMF* (represented as NS(S)MF in Figure 1.7 for brevity) provide functions to manage the LCM of network slices and network slice subnets respectively (represented by the NS(S) LCM entity in the Figure). Again, a Translator is used to map requirements from the CSMF into the underlying resources. These entities also use information provided from the FCAPS and policy frameworks to assess and decide whether network slices can be deployed or the existing ones can be reused, and to manage the correct behavior of the existing slices. After this verification, The NS(S)MF triggers the MANO framework inside the OAM block to perform LCM actions of network services that compose the network slices. MANO uses Network Service Descriptors (NSD) as a guide to compose these services using VNF. Controllers are used to rely the LCM

1. https://www.onap.org/

actions into the infrastructure. Since the OSS block has the visibility of the network slices for a communication service, it constitutes the most important part of the thesis. In consequence, the developments and contributions are going to be located in this block.

OAM

This management block contains an *FCAPS framework* which groups the fault, configuration, accounting, performance and security capabilities. It helps to know if a service is satisfying the customer requirements. To do so, it provides a monitoring scheme that covers the whole network service and its inner building blocks. Some traditional entities that perform these tasks involve:

- Element management system (EMS): performs monitoring of system configuration information e.g., tracking changes of running configuration, software versions.
- Collector: receives management information from network functions, related to alarms and threshold violations.
- Network management system (NMS): consolidates monitoring information for the network elements, network services and network slices. This entity has the visibility of all the network slices under the domain of the CSP, considering the fact that the CSP must monitor all the slices.
- QoS management: captures network element, network service and network slice performance and compliance with SLAs.
- Security: tracks security threats and events according to the traffic that is being monitored.

The usage of an FCAPS environment is crucial because it also provides management capabilities that can be exposed to the customers which constitutes added value for them. FCAPS services can be deployed using solutions from networks vendors such as HP Network Management Center [2] or Cisco Prime [3].

A *Policy framework* is used to guarantee that the business objectives of the customer are being met. Some elements that can be considered inside this framework cover:

- Policy library: contains predefined policies that can be applied to network services. These policies could make part of the major service classes detailed in Section 1.2.2. Fine grained control of the parameters of the policy can be provided.

2. https://www.hpe.com/fr/fr/networking/management.html
3. https://www.cisco.com/c/fr-fr/products/cloud-systems-management/prime-infrastructure/index.html

— Policy Decision Point (PDP): evaluates the desired behavior against the current behavior of a network service and prompts for an action to the policy enforcement point.
— Policy Enforcement Point (PEP): implements compliance with policy according to instructions received from the PDP. Each network function that is used for a network service should be considered as an enforcement point because it has the capabilities to perform corrections to parameters in order to achieve compliance with policy.

In joint effort with the FCAPS framework, a comparison can be made between what is experienced in the network against what was promised to the customer and, if there are differences, prompt for an action against the CSMF or NSMF to correct this issue.

All the previous elements exchange information with the *MANO* system. Entities that belong to this entity are borrowed from ETSI architecture. Elements contained on this framework are the NFVO, the VNF Manager (VNFM) and the Virtualised Infrastructure Manager (VIM), which are described in [45] and are called the NFV-MANO architectural framework. They provide the management of network services and its realization via VNF. They perform the management of the resources at infrastructure level that are used as foundation for their instantiation. NFV-MANO helps to solve challenges related to management of dissimilar types of resources, trust issues, and to push the usage of standardized API to manage the LCM of the network services and their resources. All of this using standardized: **(i)** Network Service Descriptor (NSD): to specify the network service; and **(ii)** VNF Descriptor (VNFD): the repository of network functions on-boarded by the CSP or uploaded by the customer.

Controllers are entities that receive orders from OAM block entities and translate them into instructions sent to the underlying infrastructure in order to realize the intended behavior. They abstract the heterogeneous nature of the infrastructure and present them as a pool of homogeneous resources to the upper layers. The target architecture should consider a controller entity that is able to provide a flexible API to applications and the ability to connect to several types of infrastructure. This could be achieved by arranging several types of controllers that interact with different infrastructure and cloud providers (e.g. Openstack, VMware, Rackspace, Azure, etc) in order to manage resources comprised of several types of technologies.

Due to the diversity of vendors and cloud providers that rent infrastructure, this controller entity would need to be scalable enough to talk to all these providers and

understand the protocols to perform its function. This way, the objective of visualizing all the infrastructure as a big pool of resources would be fulfilled.

Physical virtual resources

Resources relate to the assets that the network slice is going to use. On previous sections we stated that during the deployment process of a network slice, only the necessary resources would be used in an optimal way. In order to do so, the architecture should take advantage of key enabling technologies such as NFV, SDN and cloud computing:

- Regarding NFV, ETSI-NFV uses network service descriptors to hold the parameters to specify the network service. Its management and orchestration framework plays an important role to perform the life cycle management of the required VNF and the underlying infrastructure.
- Concerning SDN, it would provide a mechanism to offer resource abstraction to upper layers of the infrastructure. Generic API exposed via the northbound interface (NBI) will enable upper-layer management entities to send commands to the SDN controller. Via the Service Based Interface (SBI) the controller will use protocols to control the underlying resources. Via east-west bound interfaces, communication with other controllers can be established, in order to provide resiliency and create complex services.
- Regarding cloud computing, this concept enables resource sharing, allows flexibility and the creation of pools of resources. This way companies benefit from economies of scale.

Final remarks

The mentioned components provide the required functionality necessary to realize the network slicing concept. They enable the management and orchestration of the existing network slices in the infrastructure, ensuring interoperability and compliance with the required functionality by each SDO. Setting up all elements could be a difficult task, but it paves the way to the realization of the network slicing concept as it is desired. In addition, the realization of the network slice concept should consider its own security and the security of the offered service. For this, a secured-by-design and isolated-by-design network slice approach is proposed in the next subsection.

1.5.3 Secured and isolated by design approach

The secured and isolated by design architecture for network slices has two points of view: the internal to the network slice and the one that has the CSP. Both of them use two concepts, called the Virtual Security Function (VSF) and the hook.

1. The Virtual Security Function (VSF) concept: VSF is the given name to any VNF that has a function useful for security purposes. Examples of such entity are a firewall, an Intrusion Detection System (IDS), an Intrusion Prevention System (IPS).
2. The hook concept: the hook concept makes reference to the quality of a network design to reserve a place on a link to instantiate a VSF. This adds flexibility to the design allowing to modify service by adding a special functionality.

A VSF deployed on a hook helps to tighten and filter the traffic that a NF is supposed to send and receive, limiting the options for an attacker to render the service unusable or to steal information. The VSF and the dynamic hook approach take advantage of the NFV and SDN paradigms to help to increase the security of the services. They are used to enhance the security of the deployment and minimize the risks. They are used for the network slice view and the CSP view.

Network slice view

The proposed security architecture for a 5G network slice is depicted on Figure 1.8. On it, we see several VSF to guarantee security for intra-slice network functions. The entities in dotted lines have the purpose to provide high availability and reliability to the service, if the customer directly ask for it or if the service implicitly requires it.

CSP view

The CSP has visibility of all the network slices that are deployed in its infrastructure. Since it is possible that the service provider has no influence on the entities inside the network slices, there could be some security risks and threats coming from those customers. The CSP must shield from them and secure the interactions among them, as depicted in Figure 1.9.

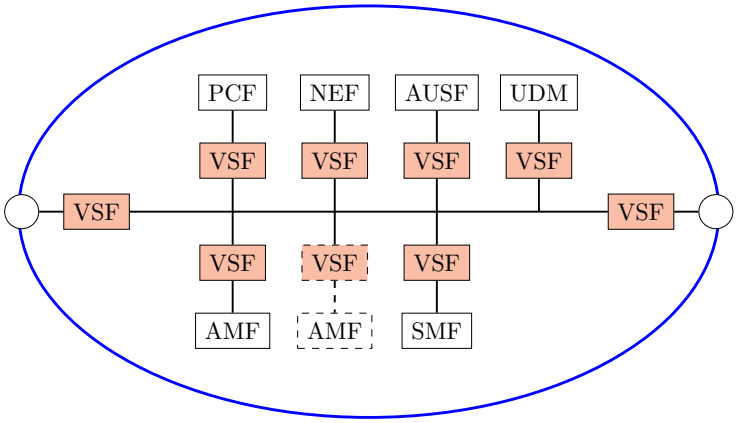

Figure 1.8 – Network slice view.

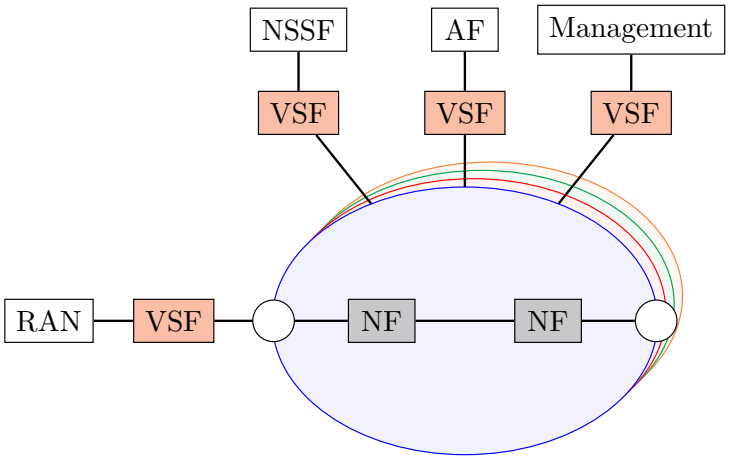

Figure 1.9 – CSP view.

1.6 Discussion

The network slicing concept must be flexible enough to cover the current use cases and the ones to come. The proposed definition in Section 1.5.1 and components specified in Section 1.5.2 guarantee the flexibility and capability to comply with the expectations of future applications for the different verticals.

The application of network slicing in several environments poses a lot of challenges, spanning topics such as orchestration, management of resources, isolation, security and the radio access network. As time goes by, the list of challenges will grow as new necessities arise. The blockage points regarding multi-domain and multi-tenant orchestration, federation and QoS decimate the intended vision of services deployed over network slicing.

There is a need to focus on certain challenges related to security. The election of

the challenges to work on was based on the novelty of the challenge, the amount of research about it (if it has little research, it is better) the degree of difficulty and the industry needs. According to these criteria, it was decided to focus on the inter-slice isolation and the management of the security for network slices in 5G Networks. The next chapters demonstrate the evolutionary process to tackle these challenges, evolving from the modeling of intra-slice interactions to then be able to control the access between network slices in a secured fashion.

Chapter 2

Formalization of a security access control model for the 5G System

2.1 Introduction

One key component that must be considered when implementing services is security. Related to 5G, there are projects that address security from the service point of view, on top of the existing services of the 5G System (5GS) (the 5GS is defined as a 3GPP system consisting of 5G AN, 5G CN and UE [17]). However, there is no clear access control model for the entities that are inside the 5GS. This fact is important, because under the future conditions and dynamic nature of the 5GS, those entities can be provided by different stakeholders.

Due to this administrative differentiation, dissimilar security levels are applied to those components. Each administrator manages their infrastructures according to their own internal rules, policies and security requirements. The need for interconnection of components poses the risk of being exposed to threats from other players, and in consequence, a secure interaction should be guaranteed to minimize the security risks. An interesting challenge is how to manage the interaction between those entities.

To tackle this problem, this chapter focuses on the access control mechanism by which a NF will have to comply in order to access another NF inside the 5GS. The objective is not to avoid all communications and completely isolate the NF from other components, it is to permit the needed communication securely via an access control mechanism. This way, we help the CSP to automatize the policy management in its own network, being the multi-provider use case scenario not considered.

There is an extensive research activity about the different access control models. The ones that are most used are Role Based Access Control (RBAC) [46], Attribute-Based Access Control (ABAC) [47], Domain and Type Enforcement (DTE) [48], and lattice-based such as the ones proposed by Bell La Padula [49] and Denning [50]. Each of them

has its own properties and mechanisms that seek to control access from subjects to objects. Their properties can be applied to several use cases, mostly in access control to documents that have different classification levels, file systems and operating systems that manage shared pools of information.

In this chapter, we will: **(i)** search for the best approach to implement secure access control inside the 5GS from the current models; **(ii)** create a security access control model that complies with the 5GS scenario requirements; and **(iii)** provide the proper mathematical definition for the model.

The chapter is organized as follows: Section 2.2 investigates how existing access control models can apply to 5GS use cases. Section 2.3 describes the components that are needed in a secure access control model for the 5GS. Section 2.4 provides the global description of the access control model. Section 2.5 describes the auxiliary concepts needed to glue its components together. Interactions inside the 5GS are presented in Section 2.6. In Section 2.7 a discussion is provided about the contents of this chapter.

2.2 Approach and architectures

Access control models constitute an area of major research due to the necessity to provide secure access to resources. In the following subsections, we will review their most important qualities.

2.2.1 RBAC

RBAC is based on the premise that the ability (or need) to access information may depend on the job function of the entity that seeks access. RBAC leverages on the **role** concept as a way to group job functions and, is based on that role that authorizes actions. The role is the mechanism to restrict the impact of the actions of a user. Besides the role, RBAC uses the concept of **users** and **permissions**, which are assigned to the roles via assignment relations, as detailed in [51]. It is interesting to consider the role as an aggregation concept, that is used to handle permissions at a higher level, easing management of the users, via their roles. Other advantages are to have hierarchies in roles, provide least privilege and separation of duties and the use of object classes as a mechanism to avoid assigning permission to specific objects [52].

2.2.2 ABAC

This model controls access to objects by making an evaluation of the attributes of entities (objects and subjects), the intended operations and the environment conditions on which resides the access request [47]. Attributes have a hierarchical structure and the inclusion of more attributes enables to have more possible rules to express policies. ABAC permits the creation of access rules without specifying individual relationships between each subject and each object. When adding new subjects, rules and objects do not need modification as subjects are assigned to the correct attributes.

2.2.3 DTE

DTE is an enhanced version of Type Enforcement, a table-oriented Mandatory Access Control (MAC) mechanism. Its improvement compared to Type Enforcement, consists in the specification of policies in a high-level language (instead of using tables) and providing implicit security attributes for objects [53]. DTE uses the concepts of **Domain** (which is an invariant access control attribute) and **Type** (which is an invariant attribute) as principal components in the model, regulating their interactions. The implementation made over the Linux kernel [48] considers that **Type** can be assigned to objects and **Domain** to processes. The DTE policy restricts access between domains and from domains to types. In this model, it is useful to ponder the domain concept, as a mechanism to provide segmentation of the resources and border authorization control point to allow the execution of actions on it.

2.2.4 Lattice-based access control

This model was developed to address the way information flows in a computer system. It mostly covers confidentiality, and also applies to integrity [54]. Under this category, we find some representative models, such as:

Bell-LaPadula (BLP)

It is a state-machine model for information flow and access control. BLP covers confidentiality only (integrity is achieved when BLP is extended by BIBA model [55]), and the secure state is permitted according to a specific security policy. This policy is summarized in three principles: simple security property; star property and strong star property, as it

is detailed in [56]. Apart of only covering confidentiality, and considering its parameters as an ordered set, it provides no native way to manage (that is, change the assignment and modification) of the classification categories. The MAC and information flow approach is interesting under this model, but its security functions only considers the security level of the subject and object.

Denning's lattice model

The most representative model under this category is the one described by Denning [50], in which she states the importance to secure information flow among Security Class (SC) in a computer system. She leverages on the use of lattices to formulate concisely the security requirements to then aid to formulate the enforcement mechanisms. The model is built over three components: **(i)** the SC, **(ii)** a flow relation on pairs of Security Classes, and **(iii)** a binary class-combining operator on SC. Using those components, Denning formulates some axioms, which are detailed in [50].

The lattice approach is exigent, and is built on a strong categorization and hierarchy. Nonetheless a more malleable approach is needed, due to the fact that our target systems would consist of a dissimilar number of objects that do not have a strict hierarchy relationship to constitute a lattice as an ordered set.

2.2.5 Access control implementations for 5G

According to the explored literature, access control implementations cover **(i)** at the application level: Internet of Things (IoT) systems, connected vehicles, medical oriented scenarios and document management; **(ii)** at the resource level: cloud scenarios, security under NFV MANO environment and traffic segmentation.

Some publications seek to apply Multi Layer Security (MLS) to telecommunication networks. For example, in [57] authors propose a modified BLP security model to be used in a 5G/IoT use case. Their approach is an evolution from the current trust model (between user and network) into a new trust model (between applications and the network). Their security model considers a scheme to label data based on the secrecy level and category, as well as capability token that rules the access scheme.

In [58] authors propose a MLS model based on BLP to avoid leakage of information from internal users in the private cloud environment, being this a key feature in order to have the ability to change the security level of an object dynamically.

In [59], authors deal with the secure distribution of workloads in the cloud by transforming the workloads, and detecting possible breaches in the inter-cloud communication. The transformation process involves awareness of the nature of the data in the workflow, the location of the clouds along with their security level.

Authors in [60] address the security in IoT in relation to the complex data flows. Even if a strict approach using Denning's lattice model can be implemented, authors prove that using a partial order model can achieve security and more flexibility.

Authors in [61] argue that authorization to access documents in a network is usually enforced at the server side. But since the network is used by actors with different clearances, malicious users can access unauthorized content by attacking the network directly.

Authors in [61] developed an Access Control Application on a SDN controller to classify the information flows and separate them by implementing VLANs, one for each group of users with similar security clearance.

In [62] authors use several access control models to allow or deny the interactions between virtual objects (VO) in the IoT environment that uses publish / subscribe mechanisms for communication. For the operational interactions, they use Access Control List (ACL), capabilities-based access control along with ABAC. They propose the use of RBAC to control the tasks performed by the administrators that configure the VOs. The policies specify which VOs are allowed to publish to which topics, and likewise which VOs can subscribe to which topics. The restricted interaction in the virtual layer maps into the physical layer.

Authors in [63] propose an Authentication and Access Control (AAC) mechanism called SSAAC (Slice Specific Authentication and Access Control), which permits the delegation of the AAC of IoT devices to the third parties that provide those devices. For this, they re-designed the Radio Access Network (RAN) architecture in order to host three additional NF that perform the AAC functions and the routing of the AAC requests to the third parties that provide those IoT devices. Their proposal maximizes the decoupling between RAN and CN; provides a way to use other mechanisms to deliver session keys to the device and the AN to secure their communication (not only use AKA); and, an attack on the AAC mechanism only affects the third party's function and the corresponding 3rd party's network slice. This leads to a decrease of the load in the CN and increases the flexibility and modularity of the 5G network. This proposal addresses a concrete IoT scenario and focuses on the RAN modifying the flows of the attach procedure. Our proposal is services independent: it seeks to manage the access between 5G entities leveraging on

the standard 3GPP procedures, taking into account security attributes and the intended actions between entities.

In [64] authors analyze the issue of confidential information carried by video signals transmitted by objects in a Vehicular Ad-Hoc Network (VANET) that uses 5G networks. In addition to cryptography to ensure secure communication, the scheme uses enhanced RBAC to allow only authorities to view video files residing in the storage system. In a similar way, authors in [65] use RBAC principles to provide assurance in access to Body Area Networks (BAN) that collect health information about a patient. Their quest is to address the situation in which a person without the required role can access the patient's data via a 5G network in order to save a life. Their proposed Emergency-aware RBAC resides on a Personal Trusted Gateway (PTG) that regulates the access to external actors into the BAN.

Authors in [66] propose to enhance the Topology and Orchestration Specification for Cloud Applications (TOSCA) modeling language with security parameters. The idea is to leverage on the SDN paradigm to use these parameters and, via an access control model, deploy services on VNF with embedded security countermeasures. Their approach conceives a security orchestrator with an access control meta model that can specify different access control models according to the needs of each tenant. This access control model approach on the TOSCA model proposes its application on 5G networks, but it does not take into account: **(i)** the inner interactions between its components according to 3GPP standards; and **(ii)** the hierarchies that are needed in order to supervise the access among those components. Due to the fact that our work resides mostly on the management functions established by 3GPP, it can address the governance of the 5G NF. These are added-value ideas in our contribution.

According to our review, most of the works are about regulating access control for the applications that run on top of the 5G network (IoT, VANET or medical environments). There is a very interesting work on an access control model for the 5GS focused on the TOSCA model and NFV MANO: even though their approach considers its application on 5G networks, it does not consider the inner interactions between its components according to 3GPP standards or the hierarchies that are needed in order to supervise the access among them.

2.2.6 Discussion

From these research works, it is inferred that the 5GS embeds a sophisticated scenario that must be controlled somehow: it has a lot of elements and complex interactions, the ecosystem involves multiple entities, providers and stakeholders with different security levels.

RBAC incorporates role as limiting concept to the operations available to a user, but it would be desirable to have more advanced attributes as ABAC. However, using ABAC requires the specification of environmental conditions, which is information that is not associated with any specific subject or object. Examples of conditions are the day of the week or the load of an entity. For our study, these conditions do not apply directly to the interactions between the entities in the 5GS. DTE provides the distinction between objects and processes, proposing the concept of domain as a restriction to limit the operations available to the subject. Nonetheless, its conception is oriented to operating systems, making difficult its implementation in other architecture by its own means. BLP is based on the security clearance and security classification in order to enforce information flow policies. The state of the system depends on few parameters, making it more restrictive when trying to apply it into other use cases. For the general case of lattice-based access control models, the need to establish ordered security classes makes it difficult to adapt to system in which labels are not necessary in a hierarchy.

It is deduced that choosing a single model is not enough to tackle the complexity to govern the secure access control of the 5GS. So the best approach is to pick the best qualities from the security models, taking into account that the chosen qualities depend on the target architecture and the properties that we would like to enforce. Next section will demonstrate the needed criteria to create an access control model for the 5GS.

2.3 New secure access control model for the 5GS

This section presents the key elements needed to build a secure access control model for the 5GS. For this, the Service Based Architecture (SBA) specified by 3GPP in [17] is used as a starting point. It identifies the principal NF that are considered to provide a 5G service with addition of management data analytics [67] and security proxy functions.

To tackle our problem, we begin with identifying commonalities on the SBA components, regarding, for example, their functions, their relationship to the AN, and the type and origin of the data that flows among them. The idea on these commonalities is to cre-

ate domains of interest, identify the roles that the NF play in the 5GS, identify subjects, objects and characteristics that help to establish a classification for them.

2.3.1 Roles

Each NF performs a function that can be categorized into roles. Roles can be used to describe the function of the NF in the 5GS. Roles are important because they help to limit the impact of the actions of the NF that has that role assigned. With this, we achieve granularity in access control based on the 5G architecture. We identify three major role categories for the NF that reside on the control plane: *Customer*, *Service* and *Governance*. On the SBA, they are represented in Figure 2.1 as colored boxes: Customer in purple, Service in red and Governance in yellow. Notice that certain elements can have more than one role, as the case of the Network Slice Selection Function (NSSF), which has service and governance functions. The reason for choosing these role categories is rooted in the

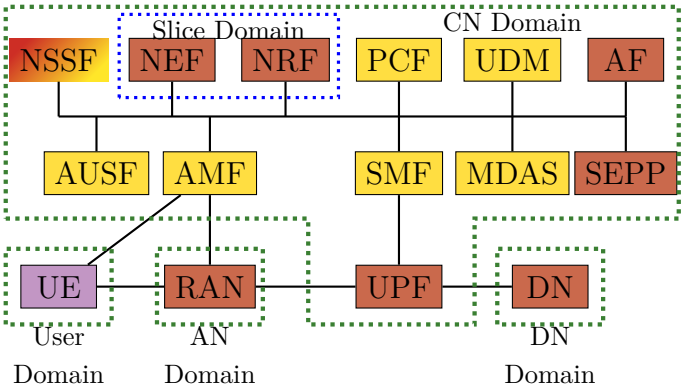

Figure 2.1 – Simplified 5G System architecture with Roles and Domains, adapted from [17].

need to grant access to the user into the network, provide a service and finally manage all the 5G system as a whole. Obviously, we can continue elaborating into the specifics for each entity, ending with the assignment of the precise NF to a concrete role. The rational for the role assignment of each NF is as follows:

1. Customer:
   (a) End user: subjects that refer to the requester of the service. It can be further divided into: **(i)** UE: when the customer relates to an equipment such as a smart-phone, with a human being controlling it. **(ii)** IoT: a connected object. This could be considered a generic use case.

(b) Communication Service Customer (CSC): refers to the entity that acts as an intermediary towards the end customer. Under this category we can find: **(i)** Mobile Virtual Network Operator (MVNO). **(ii)** Vertical industries in general.

2. Service: Figure 2.2 shows the division scheme for this role category.

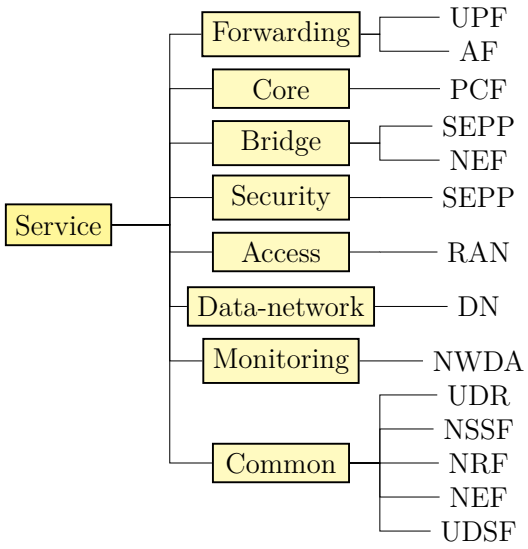

Figure 2.2 – Hierarchy for the service Role category in the SBA.

(a) Common: NF that serve a common service to subjects. For example, a VNF that is shareable and can be accessed by a VNF from other domains. NF in this category are: Unstructured Data Storage Function (UDSF), Network Exposure Function (NEF), Network Repository Function (NRF), Network Slice Selection Function (NSSF), Unified Data Repository (UDR).

(b) Monitoring: NF that have monitoring capabilities. For example, the VNF with FCAPS capabilities would monitor the service and also its constituent elements (VNF). Another NF in this category is NetWork Data Analytics Function (NWDA).

(c) Bridge: NF that are at the border between two domains, acting as translators or proxies between them. NF in this role are the SEcurity Protection Proxy (SEPP) (provides topology-hiding and intra-provider policy and filtering capabilities) and NEF (provides secure exposure of capabilities and events, translation of internal-external information towards other NF or domains) [17].

(d) Security: NF that provide generic security functionality such as firewalls, IDS, IPS or SEPP.

(e) Forwarding: NF whose function is to forward or further process traffic. NF considered under this role can be different with respect to the plane of operation, for example: an Application Function (AF) that performs a control plane task regarding traffic treatment; and the User Plane Function (UPF) which treats user traffic towards the Data Network (DN).

(f) Data-Network, Core and Access: Refer to NF whose relationship is only with other NF that are inside the same domain. This means, the NF are not exposed to the outside of the domain.

3. Governance: Figure 2.3 shows the NF role assignment to this category.

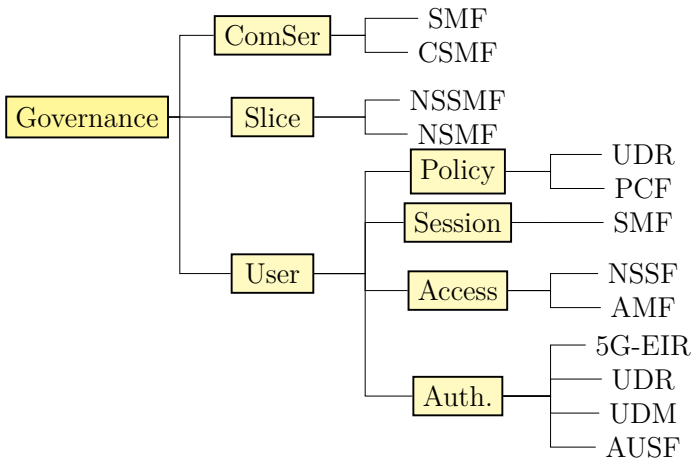

Figure 2.3 – Hierarchy for the governance role category in the SBA.

(a) User: NF that manage the user requests. Since these user management tasks comprise diverse functionalities, it is further divided into several sub-categories such as: **(i)** Authentication: secure access, authentication and authorization to the network. Roles in this category cover AUthentication Server Function (AUSF), Unified Data Management (UDM), UDR, 5G-Equipment Identity Register (5G-EIR). **(ii)** Access: NF that support 3GPP or non-3GPP [1] access modes, Non-Access Stratum (NAS) termination, mobility management, connection management. Roles in this category cover Access and Mobility Management Function (AMF) and NSSF. **(iii)** Policy: NF that deal with the policies and charging schemes for the user. Roles in this category cover Policy Control Function (PCF) and UDR. **(iv)** Session: Entities that manage the session, flows

1. WLAN is an example of non-3GPP access

or bearers for the user, according to the technology. The Session Management Function (SMF) is a role under this category.

(b) Slice: NF that perform the management of network slices. Roles in this category are the Network Slice Management Function (NSMF) and the Network Slice Subnet Management Function (NSSMF).

(c) ComSer: NF that manage the life cycle of the communication service. Roles in this category are the Communication Service Management Function (CSMF) and Session Management Function (SMF).

2.3.2 Domains

Besides a division by functionality, the 5GS can be divided into administrative areas. Usually this segmentation is called a domain. A domain is a grouping of network entities according to physical or logical aspects that are relevant for a 5G network [68]. Relevant aspects can include type of functionality, trust, (geographical) location, among others. Again, Figure 2.1 is used to show the division of the 5GS into the proposed domains.

From an *end-to-end* point of view, the well-known division of a mobile network into AN, CN and DN is reusable under this context. An example of entities located on the AN domain are wireless base stations from different technologies referred as RAN, fiber or cable modem termination systems. In the CN domain, there are network functions from 4G (like Mobility and Management Entity (MME), Serving GateWay/PDN GateWay User Plane functions (S/PGW)) and 5G (like SMF, AMF). In the DN domain, operators place functions that provide a final service functionality or interconnection to other service networks. The end-to-end service is composed of individual network slices located each one in the AN, CN, DN or a combination of those. For this reason the Slice-AN, Slice-CN, and Slice-DN domains are proposed. Specifically for the Slice-CN Domain, an important consideration is the high quantity of NF that belongs to this domain. In consequence, it is necessary to break down this domain into smaller ones. Leveraging on the 5G SBA, the proposition is to divide the CN into **(i)** sub-domains that hold NF internal to the CN, named CN-I; **(ii)** sub-domains for the NF that are exposed to other external domains, called CN-E; and **(iii)** sub-domains that have NF with governance capabilities, named CN-G.

It is necessary to consider that a communication service can be provided via a single network slice. In consequence, a *slice* domain is proposed, which constitutes a domain by

itself.

Finally, the users and industry verticals are found in a generic *Consumer* domain, for example, smart-phones, IoT devices or MVNO. Figure 2.4 provides the graphical description of the domains for the proposed security model.

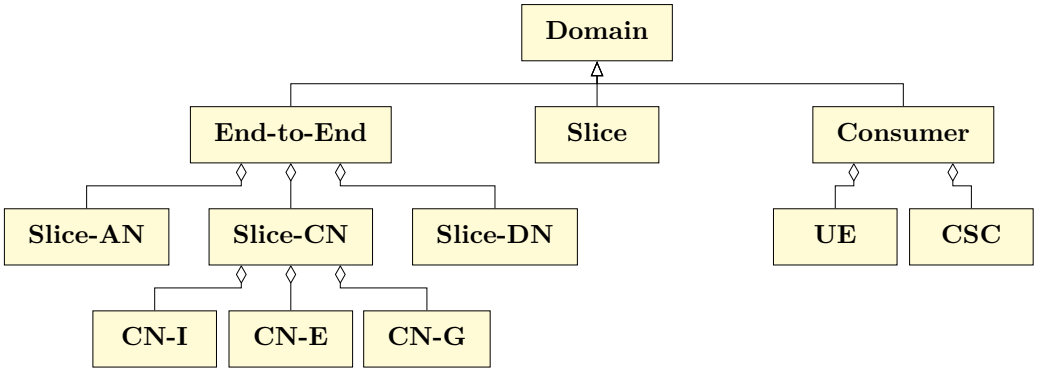

Figure 2.4 – Domain hierarchy for the proposed security model.

2.3.3 Subjects and objects

Usually, the user is seen as the human being that gets access to a resource, e.g., a file in an operating system. Under the 5G scenario, not only it is considered that a human is going to be gaining access to a service via the UE, but also a diversity of objects connected to the network. Moreover, inside the CN, we must consider the communication between NF, since their interconnection and interaction is needed in order to have the complete information to provide a service. As a consequence, the proposed model considers *subjects* as similar to a user, who performs as the active entity, initiating an interaction with a passive entity; and *objects* that offer a service and are waiting for a request, as a passive entity.

2.3.4 Other components

Besides the role, domains, subjects and objects, other components are needed to build the access control model. They are, for example, the sessions that are established by the subjects, the actions that can be performed over the objects and the permissions that are assigned to the roles in order to perform the intended actions.

2.3.5 Which model to apply?

As shown in Section 2.2 we have several models for controlling access to a system. From them, we find relevant **(i)** the **role** concept as a mechanism to efficiently manage the actions and permissions to subjects under the same function; **(ii)** the **domain** concept as a way to further confine interactions and group entities that have the same qualities, administrative management and policies; and the differentiation between **(iii) subjects** and **(iv) objects** as division of the interacting parties. Regarding their qualities, we find that it is not necessary to have a strict ordered set on the NF that conform the 5GS. These concepts lead to the choice of using RBAC and DTE models as foundation for the proposed secured access control mechanism for the 5GS. The challenge is to merge the most representative and useful components from these two models. There are similarities on the way they define role and type, users and subjects, objects and passive entities. There are concepts that are unique to each model (like the ones referring to session and domain), nonetheless the commonalities pave a way to construct a model that picks the best from RBAC and DTE, which we call, Role and Domain Access Control (RDAC).

2.4 Global description of the model

This section describes the components of the proposed model, called RDAC, which combines the best qualities of RBAC and DTE access control models. The intention is not to have a unified model with all components of RBAC and DTE, but to consolidate the required concepts according to the needs stated in Section 2.3.5. A list of abbreviations is provided in Table 2.1 to ease the reading. In the table, the mathematical definitions use a simple convention: the capital letter means that it refers to a set and the lowercase letter refers to the components of the set. For example, $\mathcal{O}$ refers to a set of objects, being $\mathcal{O}$ composed of three objects o_1, o_2, o_3. The global architecture is showed in Figure 2.5 as an UML diagram. It shows the relationships between the included components in the model: subject, session, role, domain, object, security constraint, action and permission. The UML diagram can be read as follows:

— One subject can establish one or more sessions.
— One session can activate one or more roles.
— One session originates from one domain.
— A session has security constraints.
— Objects belong to one domain and have security constraints.

Table 2.1 – Abbreviations used to describe the security access model.

Concept	**Abbreviation**	**Section**
Entity	$\mathcal{E}$, e	2.4.1
Objects	O, o	2.4.1
Subjects	SU, su	2.4.1
Roles	R, r	2.4.2
Security constraint	Φ, ϕ	2.4.3
Domains	D, d	2.4.4
Sessions	$\mathcal{SE}$, se	2.4.5
Actions	$\mathcal{ACC}$, acc	2.4.6
Permissions	$\mathcal{PER}$, per	2.4.7
Decision	$\mathcal{DE}$	2.4.7
Policies	Π, π	2.4.8
Message	$\mathcal{MES}$, mes	2.5.1
Permission to role	PerR	2.5.3
Object to domain	OD	2.5.3
Session to domain	SeD	2.5.3
Role hierarchy	RH	2.4.2
Domain Hierarchy	DH	2.4.4
Metric	m	3.5.2
Attribute	$\mathcal{A}$, a	3.5.1

— Permissions are composed of domains, actions, objects and security constraints.
— Permissions are assigned to roles.

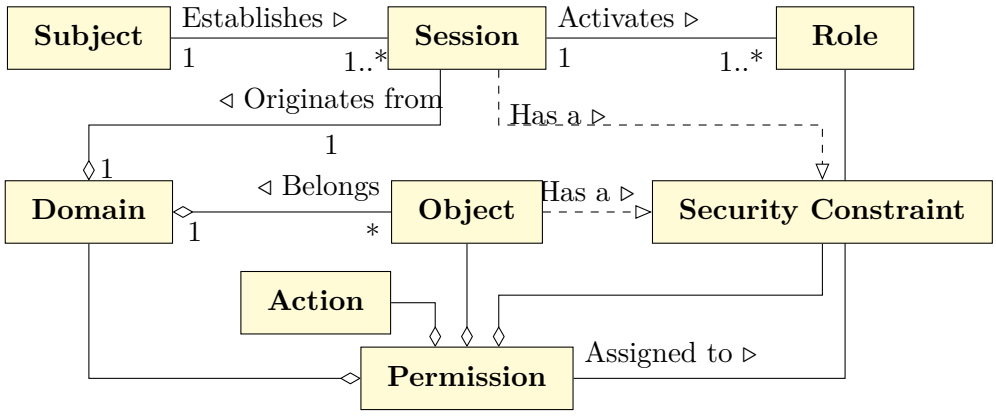

Figure 2.5 – UML representation of the RDAC model.

The components will be described in more detail in the following subsections.

2.4.1 Entities: subjects and objects

Entities denote the generic name of the actors that interact in our model. From this class, we can differentiate an entity called **subject** that denotes all the active entities. They are the ones that issue a request for a service or for information. Examples are VNF or a person as a user. **Objects** denote the entities that wait for a request from a subject. They are conceived as subjects, but in their construction they are composed of an additional standby component, that represents their ability to receive requests, to then provide a response in return. Objects represent the assets to protect. Figure 2.6 presents the concept of entity as an UML diagram. Entities can built from a finite but unbounded

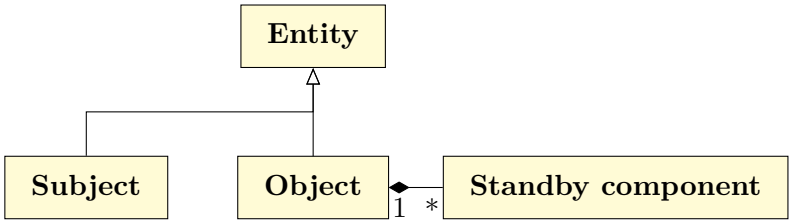

Figure 2.6 – UML representation of the entities: subject and object.

number of elements, defined as: $\mathcal{E} = \{e_1, e_2, ..., e_i\}$. Subjects $\mathcal{SU}$ and objects $\mathcal{O}$ would be represented as sets: $\mathcal{SU} = \{su_1, su_2, ..., su_j\}$; $\mathcal{O} = \{o_1, o_2, ..., o_k\}$.

2.4.2 Roles

A role is defined as "a job function in an organization that describes the authority and responsibility conferred on a user assigned to the role" [51]. As shown in Section 2.3, roles have three major divisions: customer, service and governance. As a consequence, it is specified as a set consisting of: $\mathcal{R} = \{\text{Customer, Service, Governance}\}$. Likewise, their constituting sub-roles are also part of this set. This establishes a sense of ordering among them, that is, a hierarchy. The concept of hierarchy is used to create levels of importance for the roles. The Role Hierarchy is defined as: RH $\subseteq \mathcal{R} \times \mathcal{R}$; which is a partial order in $\mathcal{R}$.

2.4.3 Security Constraint

Security constraint, denoted by Φ, refers to the "factors that impose restrictions and limitations on the system or actual limitations associated with the use of the system" [69].

Applied to the subject under consideration, a security constraint refers to the requirements that a system should comply with in relation to security parameters. Examples could be the encryption level of a Virtual Private Network, or the protocol that must be used in a communication. These requirements are the security conditions that the entities have to comply with i.e., each interaction from subject to object must guarantee a security attribute $\mathcal{A}$ (to be defined in chapter 3.5.1) superior or equal to the one specified in the rule. Considered attributes can be, for example, Affinity (Af), Trust (T) and Security Level (SL). Φ helps to link the attribute a and its value v. It is defined as follows:

$\Phi = \{(a_i, v_i) \mid a_i \in \{\text{Af, T, SL}\},\ v_i \in \mathbb{R} \wedge \forall j \in [1, |\mathcal{A}|] \backslash \{i\}\ a_j \neq a_i\}$.

This approach makes Φ extensible to any security attribute and will be useful in order to establish comparison of the security constraint between objects in the policy decision point.

2.4.4 Domain

The domain structure was presented in Section 2.3, which is defined as a set: $\mathcal{D} = \{\text{Consumer, End-to-end, Slice}\}$. The domain concept can be considered as a form of boundary that, associated with permissions, contributes to add granularity to decide what is permitted. This, depending on from which domain a session originates and the domain to which an object belongs. The domain concept considers hierarchies as a way to create levels of importance. Domain Hierarchy is defined as: $\text{DH} \subseteq \mathcal{D} \times \mathcal{D}$; which is a partial order in $\mathcal{D}$.

2.4.5 Session

Sessions constitute a mapping between a subject $\mathcal{SU}$, a domain $\mathcal{D}$, a security constraint ϕ and the activated subset of the set of roles $\mathcal{R}$ the user is assigned to [51]. One example to illustrate this concept is the *Protocol Data Unit (PDU) session*, which is represented as a bearer (in the 4G case) or as a flow (in the 5G case). A *PDU session* is a logical connection between the UE and the DN through which the UE receives services. Another example corresponds to the request-response interaction between NF in the SBA used by 5G.

In our proposed model, subjects, as active entities, would establish sessions to perform an interaction with objects. A subject can establish multiple sessions, conforming a set: $\mathcal{SE} = \{se_1, se_2, ...se_n\}$. A session is defined as $\mathcal{SE} = \mathcal{R} \times \mathcal{SU} \times \mathcal{D} \times \Phi$.

2.4.6 Actions

Actions are procedures that are used by subjects via sessions in order to perform operations on objects. Globally, the actions set $\mathcal{ACC}$ is defined as the union of the control and user plane actions: $\mathcal{ACC} = \mathcal{ACC}_{cp} \bigcup \mathcal{ACC}_{up}$.

In the 5G SBA, services are exposed by the 5G Core (5GC) NF and, in order to interact with them, some procedures are specified in the control plane [17] such as: $\mathcal{ACC}_{cp}$ = { Request, Response, Subscribe, Notify, ServiceDiscovery }.
Likewise, 3GPP specifies the support of "stateless" NF, where the "compute" resource is decoupled from the "storage" resource. To employ this, the used protocol should enable stateless operation [70]. Some UE to CN procedures do not use the SBI, nonetheless must be considered, for example, connection management, registration management and mobility management.

We can have simple user plane actions $\mathcal{ACC}_{up}$ (not covered by 3GPP) that are specified in a given time with: **(i)** the data in the fields that compose the PDU; **(ii)** features of the network slice that receives traffic; and **(iii)** nature of the source network slice from which it is permitted to receive traffic. Stateful traffic should be considered, due to the nature of most applications utilized by the users.

2.4.7 Permission

Describes the ability to perform an operation on a protected object or resource. A permission $\mathcal{PER}$ is defined in function of the role ($\mathcal{R}$), Domain ($\mathcal{D}$), Actions ($\mathcal{ACC}$), the Objects ($\mathcal{O}$), Security constraint (Φ) and Decision ($\mathcal{DE}$) as follows: $\mathcal{PER} = \mathcal{R} \times \mathcal{D} \times \mathcal{ACC} \times \mathcal{O} \times \Phi \times \mathcal{DE}$, with $\mathcal{DE}$ = {Allow, Deny}.

2.4.8 Policy

Policies, represented by Π, are defined as a set of rules. The policy contains all the access control rules that govern the interaction between subjects (via a session) and objects. This chapter deals with policies that describe the case of the 5GC delivered as a self-contained slice: its internal domain interactions (between AN, CN, DN) and interaction with the user domain. The specificity is because besides the 5GC, the exercise can

be generalized to describe interactions inside a slice for any service, e.g., IoT, connected vehicle, etc.

2.5 Auxiliary concepts for the global model

Section 2.4 described each of the components of the global access control model. In order to perform operations with them it is necessary to have tools to build relationships between those components. We define messages, assignment operations, functions and a compliance operator to do so.

2.5.1 Messages

Within the global model described in Section 2.4 the action concept was presented. It contains the global operations that can be performed without considering its implementation. It is necessary to specify the parameters of those actions. That is why we define a message $mes \in \mathcal{MES}$ that contains information such as IP address, logical ports, protocol and other information necessary to have a concrete message. In other words, the message is a subset of actions, but with the specification of parameters. Example:
a = Subscribe; a $\in \mathcal{ACC}$; mes = Subscribe (o, *event*); $mes \in \mathcal{MES}$;
Where mes describes the subscription for an *event* from an object o.

2.5.2 Assignment operations

The assignment operation relates the elements of two components of the model, that is, maps their interaction. The considered assignment relations are:

1. Object to domain assignment relation: OD $\subseteq \mathcal{O} \times \mathcal{D}$; contains all pair (o, d) for which $o \in \mathcal{O}$ and d $\in \mathcal{D}$.
2. Session to domain assignment relation: SeD $\subseteq \mathcal{SE} \times \mathcal{D}$; contains all pair (se, d) for which $se \in \mathcal{SE}$ and d $\in \mathcal{D}$.
3. Object to domain: oDom: $\mathcal{O} \rightarrow 2^{OD}$
4. Session to domain: seDom: $\mathcal{SE} \rightarrow \mathcal{D}$
5. Subject to Session: suSession: $\mathcal{SU} \rightarrow \mathcal{SE}$

2.5.3 Functions

Functions are used to perform the mapping between components. Concretely, functions are used inside the policy, in order to find information inside them. The functions are:

1. Subject ($\mathcal{D}$): returns the set of subject *su* that belongs to the domain d $\in \mathcal{D}$.
2. Object ($\mathcal{D}$): returns the set of object *o* that belongs to the domain d $\in \mathcal{D}$.
3. Object (π): returns the set of object *o* that match a policy $\pi \in \Pi$.
4. Session (su): returns the set of session $s \in \mathcal{SE}$ instantiated by a subject $su \in \mathcal{D}$.
5. Decision (π): returns the set of policy π that has an allowed permission.
6. Message (su, o): to specify the message that goes from *su* to an object *o*.
7. Procedure (mes): provides the name of the action that is contained in the message structure.
8. SessionPi(r, d, pi): returns the session *se* from the role $r \in \mathcal{R}$ and domain d $\in \mathcal{D}$ in the policy π.

2.5.4 Compliance Operator

Denoted by $\cong$, its purpose is to validate if the metrics m (to be defined in chapter 3.5.2) of a security attribute agrees with the security constraints Φ_i. It is defined as:

$\forall(a_k, v_k) \in m_{(s_i,s_j)}, \exists\ (a_p, v_p) \in \Phi_i \mid a_k = a_p \wedge v_k \geq v_p) \Leftrightarrow m_{(s_i,s_j)} \cong \Phi_i$

This means that for each set of attributes specified for a subject, it needs to exist a pair of the same name of attributes for the object. Subject and object refer to network slices, that is, the link between them that complies with the security constraint.

The $\geq$ symbol, the *greater or equal to* operator, provides a way to compare quantitatively the values of the attributes. The rationale of this operator is that for each pair in the source entity (session) security constraint, it needs to exist a pair of the same type in the destination entity (object) security constraint. This means that it is preferred to communicate with an object that has a greater or equal value in a security parameter compared to the one in the origin. This is clarified better with an example in Section 2.6.

2.5.5 Final remarks

The elements described in this section help to build the relationships with the concepts described in Section 2.4. We have all the pieces to design the properties that are required

for the inter domain interactions on the 5GS, as will be described in the next section. Table 2.2 shows the summary of the definitions for components that are a function of several elements of the model.

Table 2.2 – Summary of definitions.

Component	Definition
Session	$\mathcal{SE} = \mathcal{R} \times \mathcal{SU} \times \mathcal{D} \times \Phi$
Permission	$\mathcal{PER} = \mathcal{R} \times \mathcal{D} \times \mathcal{ACC} \times \mathcal{O} \times \Phi \times \mathcal{DE}$

2.6 Inter-domain interactions

Figure 2.7 depicts the permitted interactions between the domains in the 5GS. They constitute the properties of the security access control model. These interactions are inferred from the functional model of the 5G architecture [17] and the procedures and NF services [71] specified by 3GPP. In this section, first, the architecture is described via a

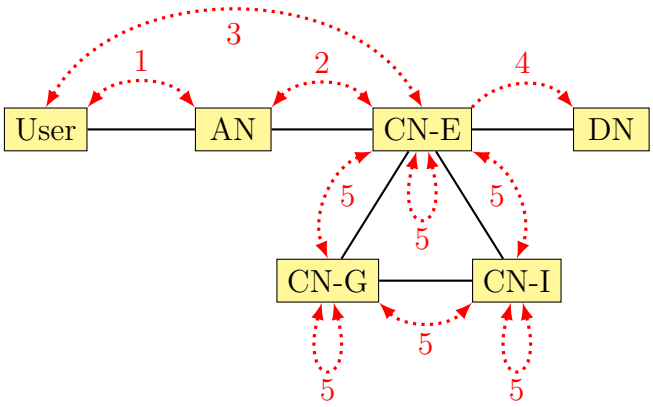

Figure 2.7 – Permitted inter-domain interaction map for the 5GS.

definition. Then, each interaction is introduced with the possible actions that can take place on it. Next, one of the interactions is taken as an example to show its properties and to demonstrate how our model is applied to it.

2.6.1 Definition of the architecture

The architecture can be characterized by Definition 2.

Definition 2. The inter-domain interactions can be represented as a graph $G = \{V, E\}$ with:
$V = \{$User, AN, CN-E, CN-G, CN-I, DN$\}$ the vertices which are the set of domains.
$E = \{$(User,AN), (AN,CN-E), (CN-E,CN-G), (CN-G,CN-I), (CN-I,CN-E), (CN-E,DN)$\}$ the set of authorized communications.

2.6.2 Interaction 1

Describes the relation between the user and AN domains. Actions that take place here are messages between UE and (R)AN during registration procedures and radio parameter exchange.

2.6.3 Interaction 2

Covers the case when NF in the AN domain need to communicate with NF in the exposed CN domain. Specifically to the RAN, considered actions are when the AMF requests the Next Generation RAN (NG-RAN) to report Radio Resource Control state information as well as NG-RAN location reporting procedures.

2.6.4 Interaction 3

Represents a typical case in the 3GPP architecture where an UE needs to have NAS communication with NF in the CN domain. For this, the AN domain acts as a transparent proxy for messages coming from the User domain, such as control plane messages or user plane traffic. The precondition is that the security constraint between User and AN domains is already active. Actions that take place here refer to procedures such as Connection, Registration and Mobility Management, mainly between the AMF and UE.

2.6.5 Interaction 4

Describes the case when traffic exits the CSP towards the DN domain. Usually this traffic is generated from entities in the User domain. There is a precondition for this traffic to traverse the NF in the AN and CN: The UE must be authenticated and with capabilities to establish a session. We assume that the security constraints in the AN and CN-E exist and are valid. Considered actions between these domains involve bidirectional user traffic such as HTTP, FTP, among others.

2.6.6 Interaction 5

This interaction, which will be used to illustrate our model, covers the case in which a NF requires communication with another NF inside the CN domain. Each CN sub-domain has different security constraints, for example, regarding the roles of the NF that belong to each domain, the functions of each domain or even whether they belong to the same stakeholder or not.

Some examples of the actions for this interaction are **(i)** the Network Function Service Framework Procedure, which includes NF service Registration, update, de-registration, NF to NF service discovery and service status subscribe/notify; **(ii)** procedures and flows for Policy Framework (when AF are involved); and **(iii)** interactions for network slice selection and communication between CSMF and NSMF.

Property statement

The inter-domain communications are allowed only for links l belonging to E, that is, the set of authorized communications, as shown in Property 1:

Property 1. $\forall\ l = (v_i, v_j) \in E \wedge \forall\ su \in Subject(v_i) \wedge \forall\ se \in Session(su)\ \wedge\ \forall\ o \in Object(v_j) \wedge \forall\, mes \in Message(s, o) \wedge \forall\, r \in Role(s) \Rightarrow \exists\, \pi \in \Pi \mid se \in SessionPi(r, d, \pi) \wedge o \in Object(\pi) \wedge \Phi(s) \cong \Phi(\pi) \wedge Procedure(mes) \subset Action(\pi) \wedge Decision(\pi) = \text{Allow}$

In addition, there must be no other communications allowed for any link composed of two different domains belonging to V and not belonging to E, as shown in Property 2:

Property 2. $\forall\ v_i, v_j \in V\ \wedge\ \forall\ su \in Subject(v_i)\ \wedge\ \forall\ se \in Session(su)\ \wedge\ \forall\ o \in Object(v_j)\ \wedge\ \forall\ mes \in Message(s, o)\ \wedge\ \forall\ r \in Role(s) \mid v_i \neq v_j\ \wedge\ l = (v_i, v_j) \nexists\ E \Rightarrow \exists\, \pi \in\ \Pi \mid se \in SessionPi(r, d, \pi) \wedge o \in Object(\pi) \wedge \Phi(s) \cong \Phi(\pi) \wedge Procedure(mes) \subset Action(\pi)\ \wedge\ Decision(\pi) = \text{Deny}$

Topology

To visualize the example, lets consider three VNF as specified in Table 2.3. The topology is shown in Figure 2.8.

A VNF with a CSMF role must create network slices by sending its configuration commands to any available VNF with a NSMF role under its command. Those NSMF have different security constraints specified by a security level SL.

Table 2.3 – Components used for the example.

Concept	Subject	Object	Object
Name	CSMF	NSMF1	NSMF2
Role	Governance→ ComSer→CSMF	-	-
Domain	End-to-end→ Slice-CN→ CN-E	End-to-end→ Slice-CN→ CN-I	End-to-end→ Slice-CN→ CN-I
Φ	(SL, high)	(SL, medium)	(SL, high)

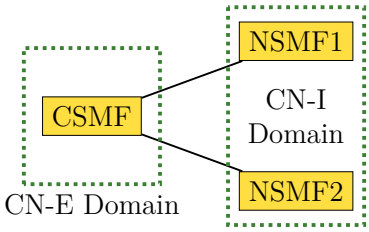

Figure 2.8 – Topology of the example for interaction 5.

Actions

Available actions $acc \in \mathcal{ACC}$ in the control plane for the governance role are: **(i)** CreateNSRequest; **(ii)** CreateNSResponse; and **(iii)** DeactivateNS; in order to create or deactivate a NS and issue a response about its creation.

Messages

Accordingly, messages $mes \in \mathcal{MES}$ are specified as **(i)** CreateNSRequest (NSD, NSParam, o); **(ii)** CreateNSResponse (NSInstanceId, o); **(iii)** DeactivateNS (NSInstanceId, o); which specify the NS description and the parameters for its creation, being the ID of the NS the return value after the success of the command.

Permissions

As a reminder, permissions are defined as $\mathcal{PER} = \mathcal{R} \times \mathcal{D} \times \mathcal{ACC} \times \mathcal{O} \times \Phi \times \mathcal{DE}$. For the current example they are specified as follows:

$\mathcal{PER}$ = { (Governance-ComSer-CSMF, CN-E, CreateNSRequest, NSMF1, (SL, medium), Allow), (Governance-ComSer-CSMF, CN-E, CreateNSRequest, NSMF2, (SL, high), Allow), (Governance-ComSer-CSMF, CN-E, DeactivateNS, NSMF1, (SL, medium), Allow), (Governance-ComSer-CSMF, CN-E, DeactivateNS, NSMF2, (SL, high), Allow) }

Assignments and functions

Thanks to assignment operations specified in Section 2.5.2 and the functions specified in Section 2.5.3 it is obtained that: **(i)** the VNF are assigned to the corresponding domains via oDom(); **(i)** we know from which domain a session originates, via seDom().

Property application

We assume that the VNF1 is able to create a session with the CSMF role and already has knowledge of the two VNF that are in the CN-I domain that serve as NSMF.

The policy requires that the actions that the session with CSMF role creates are allowed and that are permitted towards both NSMF. In the same way, the parameters of the procedure, such as the destination object and the parameters for the NS are valid and satisfy the policy.

The crucial part results from the verification of the compliance between the security constraint of the CSMF and both NSMF. For simplicity, we use only one parameter: the security level (SL). Its values are specified for all the involved entities. Since the SL for the CSMF is high, the interaction towards the NSMF1 would be denied, because $\Phi(s) \ncong \Phi(o)$ as stated in the policy. The reason is that the SL of NSMF1 is lower than the one for the CSMF. The interaction between the CSMF and NSMF2 will be allowed since the SL values are equal. This way, the property is respected according to the parameters of the SL established in the policy for the destination objects.

2.6.7 Summary

The described properties of the model allow only a certain type of communication between the entities of the desired domains. These interactions obey the specifications of 3GPP regarding the procedures that a subject must follow in order to gain access to a service in the 5GS. As done with interaction 5, the same approach can be developed for the other remaining communication scenarios described in this section.

By specifying more parameters in the security constraint and the actions in the policy, a high control is achieved, in function of the assigned permissions and the domains involved in the communication.

2.7 Discussion

This chapter addresses a major security concern for CSP when faced with resource sharing, being this technique important when deploying 5G services and expanding network coverage.

Each stakeholder has its own internal security policies and security levels, so it is necessary to establish access control mechanisms that incorporate the required elements to restrict and authorize interactions according to those constraints.

Due to the fact that traditional access control models do not fulfill the requirements of the 5G architecture, a new method was created called RDAC. This novel approach picks the best concepts of RBAC and DTE access control models.

With the concept of **role** we restrict actions according to the function of a NF. With the concept of **domain**, we restrict interactions according to the section of the 5G system or stakeholder and whether the **session** created by a *subject* has authorization to establish communication with an **object** in the destination domain. With the objective to bind the aforementioned requirements, the concept of **security constraint** was created as a mechanism to specify several security properties.

The proposed **actions**, which leverage on the functional model of the 5GS, specify the appropriate procedures that can be executed over objects, being **permissions** the concept that links these two concepts together. The *property statement* represents the description of the required allowed communication. Finally, the **compliance operator** is used as a mean to evaluate if the interaction can be authorized using the involved security constraints.

Something interesting about the proposed access control model is its extensibility: several security properties can be specified according to the needs of the CSP. Moreover, the concepts that are used are general enough to apply to other use cases and architectures.

The usage of this new model lays the foundation for secure resource sharing among the different players involved for providing services over 5G networks. As a consequence, it invites to consider departing from the intra-slice interactions case and begin to establish how to manage inter-slice interactions, as is covered in the next Chapter.

Chapter 3

Managing secure inter-slice communication in 5G Network Slice Chains

3.1 Introduction

Network slicing is one of the key enablers for the use cases that are proposed for 5G [15]. Along with SDN, NFV and cloud computing, it provides a novel partitioning scheme to instantiate a CS on top of network slices. It will use resources that belong to the same CSP that offers the service or to different operators, organizations and stakeholders [10].

Network slicing has the power to host several services over the same infrastructure, enabled by an intelligent resource sharing. Because of that shared infrastructure, security considerations must be taken into account in order to guarantee that the provided services comply with baselines established by a security policy [72].

Interactions between network slices will become commonplace, because the CSP can provide common functions through a network slice that is accessible for consumption by other dedicated slices. As network slice interconnection brings the risk of exposure to threats from other players, a secure interaction should be guaranteed to minimize security risks. In order to do so, the CSP has to set up different rules and measures according to a policy to guarantee secure inter-slice communication. This is a difficult task because network slices have different security levels, evaluated via attributes that are diverse in nature and purpose.

The evaluation of these policies carries a significant load since rich added-value services will be provided using several interconnected network slices. Policy based network management is time-intensive, complex, and expensive [73]. In consequence, not only the security requirements of the network slice need to be evaluated, but also the ones that control the access and communication between them. Since the amount of network slices

under the domain of a CSP will rise exponentially as new use cases arise, the evaluation of the security policy will become more difficult and time-consuming.

A communication service is conceived with an end-to-end scope via network slices. For this, the CS can be composed of two or more successive network slices (a chain of network slices) spanning through AN, CN and DN. Each network slice that belongs to the chain performs a specific function.

From this scenario, interesting challenges are: how to manage the interactions between network slices when each one has different security attributes and different security requirements? How to bring this to a next level when a chain of network slices is considered, in order to provide an optimal path for an enriched communication service according to the security requirements expressed by the customer? Finally, how the CSP can choose a network slice chain that uses the least security resources in order to not only comply with a constraint expressed by the customer, but also have savings in asset utilization in its infrastructure?

According to our research, as will be presented in Section 3.2, no work has been made regarding the formal model of a communication service that uses network slices, taking into account their inherent security attributes. Moreover, no study about the evaluation of these attributes when inter-slice communication is considered, specially in the case where successive network slices need to be connected. In addition, most of recent work focuses on the placement of the services on the resources. The challenge that is detected is not about provisioning resources, but about choosing a path to interconnect existing provisioned resources according to security constraints. In our case, the resources are the already deployed network slice instances.

The presented new concepts contribute to go beyond the access control models that already exist (which are more focused on the user or the resources), by adding an end-to-end view of the communication service considering the security needs for its deployment and the security specifications of the already deployed network slices.

The objectives of this chapter are summarized as: **(i)** model the network slicing structure mathematically using graph theory, leveraging on the definitions given by Standard Developing Organizations; **(ii)** deduce a general concept called network slice chain, which describes the sequence of network slices data must flow through in order to provide a communication service; **(iii)** provide properties and policy rules to validate that a network slice chain exists and can be used according to security constraints: and **(iv)** identify the path that offers minimal resource utilization for the CSP considering the security

constraints that are specified in the policy.

The chapter is organized as follows: Section 3.2 presents works related to inter-slice communication and path selection solutions to elect the best route. Section 3.3 presents an example of a common network slice set-up from a CSP, who will experience challenges regarding the secure composition of a rich communication service. Section 3.4 describes the mathematical model, definitions and properties of a network slice and network slice chain. Section 3.5 describes the different components used on the communication model. Section 3.6 describes the rules and policy validation steps that govern communication in the network slice chain, to then describe the algorithm and the error evaluation used to solve the challenges in Section 3.7. These concepts are then applied to a test-bed scenario in Section 3.8. Section 3.9 draws concluding remarks about this chapter.

3.2 Recent works

The field of *inter-slice access control* has attracted few research works. In [68], the 5G-ENSURE project focuses on the access control from end-users to the resources offered by a network slice in a 5G network. They provide a set of countermeasures and enablers for this purpose. The inter-slice communication and access control are not addressed.

In a different perspective, authors in [74] address the inter-slice communication regarding the need to guarantee isolation. They point out that improper inter-slice isolation leads to threats in network slicing. They include the suggestion to use a fine-grained access control to limit access from a tenant to the entire infrastructure.

In [75] the 5G-MoNArch project works on providing end-to-end slicing support via enablers pertaining to inter-slice control and management, which are some of the proposed innovations of their work in order to provide slice admission control. The inter-slice management still resides into the NSMF by incorporating a cross-slice management and orchestration function. With the aid of a security monitoring manager, it can manage security requirements and the establishment of security trust zones. Inter-slice management also assures that the resources assigned to the network slice instance are optimal, used wisely, at the same time guaranteeing SLA. Slice management decisions are supported by context information and an enhanced NWDA. In the same fashion, authors in [76] propose an inter-slice management mechanism to control events in a 5G network. Using queue and graph theory, they create a reference model that captures events from the network and according to their importance or impact on metrics, classifies the events for

resolution, avoiding network congestion. The projects do not provide information about access control mechanisms.

Authors in [77] present how authentication and authorization were integrated in the SONATA Service Platform, in order to manage the authentication, identity management and authorization of users and micro-services in a 5G network. Their approach is generic, supplying these security features for the users and the networks functions inside 5G. The slice use case is not mentioned, neither inter-slice communication management.

In [66] authors propose to enhance the TOSCA modeling language with security parameters. They leverage on the SDN paradigm to use these parameters and, via an access control model, deploy services on VNF with embedded security countermeasures. Their work focuses on a security orchestrator that extends the NFVO to perform the LCM of network services. This helps to enforce access control policies per tenant of their resources. Since the impact of their work is at the network service level, it is necessary to consider the use of a NSMF and CSMF in order to have visibility at network slice level and achieve inter-slice management. In a similar fashion in [78] authors propose an enhancement of the TOSCA language to model the protection of clouds, represented as resources in unikernel system instantiated in virtual machines. Again, both approaches provide a way to specify and build secured network functions, nonetheless, their approach can be improved by considering a top layer approach from a communication service point of view and considering a chain of those VNF of kernels in order to build richer services.

On the subject of service chaining, there are works that address *path selection* according to parameters such as congestion or by the usage of several algorithms to elect a best route.

In [79] authors propose a Software Defined Optical Networks Slicing Architecture that leverages on the advantages of multi-protocol label switching and SDN to route traffic between slices in optical networks. The path provisioning stage does not consider security parameters to build the path of nodes that constitute the slice. Their approach does not consider the case of consecutive slices: this means it reviews the case of two slices that need to exchange data.

In [80] authors present a scheme to dynamically segment the physical infrastructure of a RAN into network slices with different SLA levels. The slices must obey a specific resource allocation policy, which are specified by flows and the characteristics of traffic specified by traffic rules. Their contribution covers the RAN, and even though their approach considers flows into one slice, there is no consideration of security attributes or

flows traversing several network slices.

In [81] authors address a type of virtual network embedding (VNE) problem called 5G network slice provisioning. It is the process of allocating physical resources to slice requests. For it, the nodes are ranked from the perspective of multi-attribute decision making when provisioning slice nodes. In order to ensure the security of a 5G core network slice, the slice tenant may request the slice with a specific security requirement. Their approach cover the allocation of the slice requirements into the infrastructure, but no connection between slices is considered.

In [82] authors state that Over The Top (OTT) service providers should be able to modify the slices over which their services are provided, including service chain modification. Among those actions, flow prioritization could be performed, and this jeopardizes fairness with respect to other tenants. Their approach covers a flow prioritization algorithm for network service chaining, which improves OTT applications' service levels, achieving more efficient resource management. Authors mention the importance of security of the flows, but no attributes are mentioned, or used in the construction of the chain.

In [83] authors leverage on SDN qualities to permit traffic between network nodes according to security constraints. These policies are enforced by several multi-domain SDN controllers. Their approach does not cover the consideration of network slices and specifies a single security attribute, linked to the characteristics of the SDN controller.

Authors in [84] focus on finding the path that provides the best bandwidth, for the concrete case of high performance computing networking. This way, congestion is avoided by assessing the state of the link at each hop. No central entity controls the path construction, so each decision is made locally thanks to metadata received by each node, describing the conditions of the link and saturation cost for it. Their approach does not consider security attributes and does not consider the end-to-end view, which is important when considering a complete end-to-end service as is envisioned in telecommunication applications.

In [85], authors address a concrete use case scenario for an emergency vehicle that must find the route to the nearest hospital in the shortest time. For it, they use Dijkstra's algorithm to determine the shortest time to the nearest hospital and the Floyd-Warshall algorithm to know the closest distance to the hospital. The collaboration between both algorithms constitutes an improvement in order to reach a more concrete decision and can be applied to other fields. This approach is inspiring to be applied in the case where

the vehicle and the building provide services in different slices and collaboration between them is needed. In this case, for fast localization purposes.

The aforementioned works point out challenges, focus on the isolation problem, on how an end user or tenant accesses to resources, how to perform the inter-slice management and orchestration via a broker mechanism [10] and how to map customer requirements into the infrastructure according to its capabilities.

First, no formal model of the network slice environment is provided, neither security considerations for inter-slice communication when several network slices need to be connected to provide composite services. This is a central issue for a CSP that is deploying services via network slices.

Second, customer requirement mapping considers parameters such as performance, quality of service or load. Even though there is a conscience that security is important, attributes that relate with this subject are not considered explicitly on this mapping.

Moreover, these works do not consider the scenario of interest: the inter-connectivity of consecutive network slices to create communication services, considering the trade-off between the customer requirements and the security attributes of the already existing network slices and links. In addition, the optimization of the process to validate the secured connectivity of network slices and the resource usage is not taken into account.

Finally, the cited works lack of an end-to-end view and the consideration of security parameters in order to choose a path. In addition, there is no consideration of a policy seeking to enforce security for the connection according to constraints expressed by the customer.

For a better understanding an example is given in the next Section. Then, the key elements to solve these challenges will be developed. The structure of this process is shown in Figure 3.1

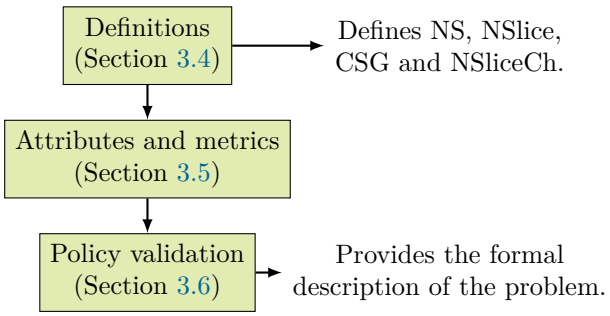

Figure 3.1 – Flow chart summarizing the needed elements to solve the challenges.

3.3 Motivating example

In order to understand better the properties and different elements that are inside the proposed model, a use-case scenario is presented. Even though it does not depict a specific service, it is generic enough to fit into any communication service offered by a CSP. The architecture is presented in Figure 3.2, which contains a set of eleven network slices, ranging from s_1 to s_{11}, connected arbitrarily according to the needs of the CSP. Where convenient in the text and with the purpose of improving readability, the abbreviation *NSlice* will be used to name a network slice. Each NSlice is configured according to a *service*

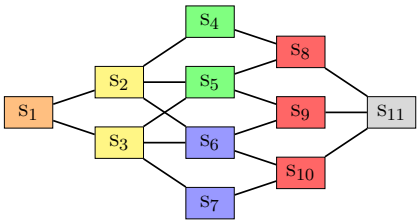

Figure 3.2 – Topology of the set of NSlice and corresponding types for a CSP.
For this example, a service starts on NSlice s_1 and arrives to NSlice s_{11}.

type to perform a specific function, as specified by 3GPP [17]: Enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low Latency Communications (URLLC), Massive Machine Type Communications (mMTC), and Vehicle to Everything (V2X). However, the provider can offer other non-standardized *service types* according to the needs. These service types are represented by a different color: orange, yellow, green, blue, red and gray. A single *service type* can be assigned to different colors, symbolizing the different configuration parameters that are used according to the needs of that concrete service.

For example, the orange NSlice could be an aggregation service NSlice for an enterprise; the yellow NSlice an IoT NSlice; green and blue NSlice constitute added value services (built from network functions to provide services such as traffic filtering, IDS/IPS); the red NSlice a 5G NSlice to provide final connectivity; and the gray NSlice a data network that provides a concrete service. A more concrete use case illustrating a similar setup is provided in Section 3.7.3.

All network slices are connected together in an ordered sequence to provide a service. For example, assume the presence of a communication service that we name CS_1. It considers the orange, yellow, green, red and gray service types. Similarly, another communication service named CS_2 has orange, yellow, blue, red and gray service types.

Each communication service CS_1 and CS_2 can be set up according to the needs from the CSP creating a Communication Service Graph (CSG). For example, regarding CS_1 it can be given at least two CSG: CSG_{11} (Figure 3.3) and CSG_{12} (Figure 3.4). The key message is that, even tough the nature of the CS is the same, each NSlice can have a different configuration and different resources, enabling to provide options of deployment according to the needs. The same approach can be made with CS_2, in which two CSG

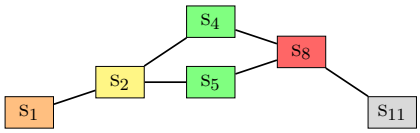

Figure 3.3 – Communication service graphs for CS_1: CSG_{11}.

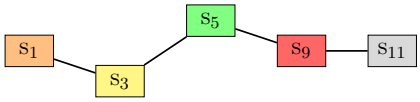

Figure 3.4 – Communication service graphs for CS_1: CSG_{12}.

are presented: CSG_{21} (Figure 3.5) and CSG_{22} (Figure 3.6). Other arrangements of CSG can be made, enriching the exercise. The advantage of considering the service as a CSG is that the operators can configure the routing of the system in order to forward the traffic through the slices according to a certain policy. With this, traffic can exploit the characteristics of the network topology and then be treated according to the specification of each NSlice. The traffic will follow a chain of slices that comply with a use-case for the customer. Concretely for CSG_{11}, the provider can set up two NSlice chains specified by

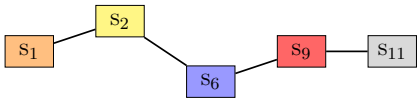

Figure 3.5 – Communication service graphs for CS_2: CSG_{21}.

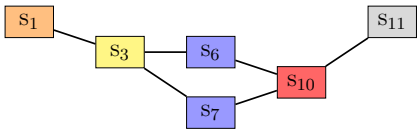

Figure 3.6 – Communication service graphs for CS_2: CSG_{22}.

blue and red arrows. As presented in Figure 3.7, the blue NSlice Chain (going through

s_1, s_2, s_4, s_8 and s_{11}) covers a green NSlice with an IDS that detects a certain type of traffic. Similarly, the red NSlice Chain (going through s_1, s_2, s_5, s_8 and s_{11}) can contain a green NSlice that has an IDS with a different detection policy. The same approach can be made for CSG_{22}, as shown in Figure 3.8. The presented topology is complex even

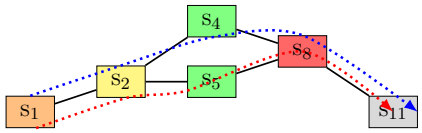

Figure 3.7 – Example of two CSG with two network slice chains: CSG_{11}.

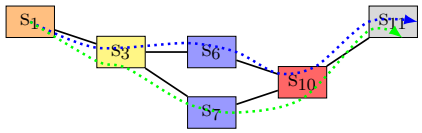

Figure 3.8 – Example of two CSG with two network slice chains: CSG_{22}.

though the number of network slices is small. As the number of network slices increases, their management becomes a challenge. This manageability has to do with the way to connect the network slices (must ensure the proper authentication and security between them), how to keep the guarantees of the service offerings to the customer and how much of the resources are going to be used to offer the CS securely. This implies that the set-up and configuration of the communication service must follow certain rules and constraints expressed in the policy, which specifies its security requirements and the type of traffic that is allowed to flow. Moreover, as the network is a dynamic entity, the topology or the service and security requirements can change, so the CSP must perform validation that a path, represented by a NSlice Chain, can be used for the service required by the customer. Not addressing the needs regarding management and security validation, makes difficult the secure deployment of rich communication services using several connected network slices.

With this setup, the next Section elaborates on the properties of NSlice and NSlice Chains, whose specification constitutes one of the major contribution of this work. On top of these foundations, the evaluation of the paths and minimal resource utilization problem will be addressed.

3.4 Network Slice and Network Slice Chain

This Section provides the mathematical background to describe the novel concept called Network Slice Chains. To do so, it evolves from its basic building blocks to then state its key properties.

3.4.1 Network Service (NS)

The network slicing model relies for its realization on the ETSI NFV concept of NS, detailed in [21]. A NS is a composition of Network Functions (NF) arranged as a set of functions with either unspecified connectivity between them or connectivity specified according to one or more forwarding graphs [86]. It is deduced that key components of a NS are VNF, Virtual Link (VL), and VNF Forwarding Graphs (VNFFG). All these elements provide a specific functionality and resource requirements for network slices, which will be presented in the next subsection.

3.4.2 Network Slice (NSlice)

3GPP [87] defines that a CS is offered by a set of network slices, being each NSlice composed by an ordered set of NS. This notion of "interconnection" leads us to represent the NSlice as a connected graph.

Definition 3. The NSlice is a graph composed of: **(i)** a non-empty set of vertices (V), which are the NS; and **(ii)** a set of edges (E), which are the VL. For a network slice $\mathcal{A}$: $\text{NSlice}_{\mathcal{A}} = (\text{NS}(\mathcal{A}), \text{VL}(\mathcal{A}))$ □

Property 3. Let $\mathcal{S}$ be the set of NSlice that belong to a CSP. $\mathcal{S} = \{ s_1, s_2, ..., s_i, ..., s_m\}$. The CSP uses s_i to provide a service to its customers and their disposition obeys the CSP's internal rules and policies. □

Property 4. Each NSlice has one *service type* that describes its function. The set of types is called $\mathcal{T}_{\mathcal{S}}$. $|\mathcal{T}_{\mathcal{S}}|$ represents the number of *service types* that are provided by the CSP. They refer to the Service Classes, as defined by 3GPP [10]. □

Property 5. The function **type** is used to know the Service Class that has the NSlice. A NSlice can have only one type. The function **type** is defined as **type**: $\mathcal{S} \rightarrow \mathcal{T}_{\mathcal{S}}$. □

The CSP uses several interconnected network slices to provide a complete service to the customer: this constitutes what is called a communication service. Next subsection formalizes this concept and the inference of the communication service graph.

3.4.3 Communication Service Graph (CSG)

A CS is defined as an ordered set of types of network slices, whose services are offered to different market segments, obeying a business purpose [19]. These network slices are connected via Network Slice Links (NSL).

Definition 4. A type of CS is defined as $\mathcal{T}_{\mathcal{CS}} = \langle \mathcal{T}_{\mathcal{CS}_1}, \mathcal{T}_{\mathcal{CS}_2}, ..., \mathcal{T}_{\mathcal{CS}_m} \rangle$, i.e., the traffic of a CS is going to flow through an ordered set of NSlices $\mathcal{T}_{\mathcal{CS}} = \langle \mathcal{T}_{\mathcal{CS}_1}, \mathcal{T}_{\mathcal{CS}_2}, ..., \mathcal{T}_{\mathcal{CS}_k} \rangle \mid \forall i \in [1; k], \mathcal{T}_{\mathcal{CS}_i} \in \mathcal{T}_{\mathcal{S}}$. □

There can exist several NSlices deployed by a CSP for a type $\mathcal{T}_{\mathcal{CS}_i}$. In fact, there exist a set $\mathcal{S}_{\mathcal{T}_{\mathcal{CS}_i}} = \{s \mid \textbf{type(s)} = \mathcal{T}_{\mathcal{CS}_i} \wedge s \in \mathcal{S}\}$. The interconnection of successive elements that belong to $\mathcal{S}_{\mathcal{T}_{\mathcal{CS}_i}}$, $\mathcal{S}_{\mathcal{T}_{\mathcal{CS}_{i+1}}}$ creates an ordered graph.

Definition 5. For a communication service CS, the CSG $\mathcal{C}$ is a directed weighted graph such as: $\mathcal{C} = (\ \mathcal{S}'$, NSL) where: $\mathcal{S}' = \{s | s \in \mathcal{S} \wedge \textbf{type(s)} \in \mathcal{T}_{\mathcal{CS}}\}$ and NSL $= \{(u, v) | u, v \in \mathcal{S} \wedge u \neq v\}$. □

Property 6. Each link $(u, v) \in$ NSL has a set of attributes $\{\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_m\}$. (u, v) inherits a quality from graph theory called weight $\mathcal{W}_{(u,v)}$ that is a function which, using the values of the attributes, computes an unified metric for (u, v). $\mathcal{W}_{(u,v)} = \mathcal{F}(\mathrm{a}_1, \mathrm{a}_2, ..., \mathrm{a}_m)$. The definition of $\mathcal{F}$ and the presentation of the attributes are explained in Section 3.5.1. □

These aforementioned definitions and properties help to define a CSG, which provides a way to deploy a concrete communication service and permit the flow of data among a subset to those network slices. That is where the concept of NSlice Chain comes to play, as is shown in the next subsection.

3.4.4 Network Slice Chain (NSliceCh)

The Network Slice Chain (NSliceCh) is conceived as a concrete path in the CSG that a flow of data follows, which complies with certain requirements related to the communication service purpose, the nature of the traffic and security attributes. The NSliceCh leverages on Definition 5, which defines the CSG as a set of NSlices whose type respects $\mathcal{T}_{\mathcal{CS}}$ over which the traffic will flow. For readability of the definition, $\mathcal{P}$ represents a NSliceCh.

Definition 6. The CSG $\mathcal{C} = (\ \mathcal{S}'$, NSL) contains a set of network slice Chains $\mathcal{P}_\mathcal{C}$, which comply with the sequence of types of Network slices $\mathcal{T}_{CS}$ and do not form a loop.
$\mathcal{P}_\mathcal{C} = \{\langle s_1, ...s_i, ...s_m\rangle \mid \forall\ i \in [1, m], s_i \in \mathcal{S}' \wedge\ \mathbf{type}(s_i) = \mathcal{T}_{\mathcal{C}\mathcal{S}_i} \wedge \nexists\ s_i \in \langle s_{i+1}, ..., s_m\rangle\}$ □

As an illustration, Figure 3.7 shows two different NSliceCh: one in red and the other in blue dotted line. It is supposed that they comply with the demands from the CS and its security constraints. With all the previous definitions, all the elements are provided in order to use the tools to assess inter-slice communication.

3.5 Attributes and metrics involved in inter-slice communication

From the mathematical representations, definitions and properties shown in Section 3.4, we define the attributes and metrics needed to manage inter-slice communication. Specifically, these are the *attributes* of network slices and their corresponding measurement using *metrics*.

3.5.1 Attributes

Attributes were first mentioned in Section 2.4.3. Attributes are an abstraction that refer to a characteristic or property of an entity that are useful for the implementation of access control and flow control policies [88]. The attributes to be selected depend on the business or support functions that want to be enforced by the CSP. In this thesis, the selected attributes play a role in providing isolation and they give information to assess whether the connection between network slices can be permitted or not. For this, attributes that are important from a security perspective are Affinity (Af), Trust (T) and Security Level (SL). In this subsection, operations are proposed among them, being these operations a particular case of the function $\mathcal{F}$ stated in Property 6.

Let $\mathcal{C} = (\ \mathcal{S}'$, NSL). Each NSlice $s \in \mathcal{S}'$ has a set of attributes defined as $\mathcal{A}(s) = \{(a_i, v_i) \mid a_i \in \{\text{Af, T, SL}\}, v_i \in \mathbb{R} \wedge \forall j \in [1, |\mathcal{A}|]\backslash\{i\}\ a_j \neq a_i\}$.

Each attribute is defined and specified according to formulas and properties as follows:

Affinity (Af):

It is used to avoid conflicts regarding the nature of the offered slices, helping to determine whether they can be connected or can coexist.

Affinity has a nominal type of data, specified by the network administrator. Considered values are the basic service classes for 5G established by 3GPP [10]. However, the CSP is free to provider additional service classes according to the need, for example, it can use a *common* service type that contains regular functionality and aids to connect dissimilar NSlice.

Property 7. Affinity for a link $(s_i, s_j) \in$ NSL is achieved if the (s_i, s_j) that make it up have the same affinity parameter. We call $\mathcal{F}_{\text{Af}}$ the function that finds the affinity for a link $(s_i, s_j) \in$ NSL.

$\mathcal{F}_{\text{Af}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$

$\forall s_i, s_j \in \mathcal{S}', \exists\ (a_{i_p}, v_{i_p}) \in \mathcal{A}(s_i) \wedge (a_{j_k}, v_{j_k}) \in \mathcal{A}(s_j) \mid a_{i_p} = a_{j_k} = \text{Af} \Rightarrow$

$$\mathcal{F}_{\text{Af}}(s_i, s_j) = \begin{cases} 1, \text{if } v_{ip} = v_{jk} \\ 0, \text{otherwise} \end{cases}$$

This means that if the services belong to the same service class, their affinities are the same and the function will have 1 as a result. □

Property 8. Affinity for a NSliceCh $\mathcal{P}$:

Let $\mathcal{C} = (\ \mathcal{S}'$, NSL). $\forall\ \mathcal{P} = \langle s_1, s_2, ..., s_n \rangle \in \mathcal{P}_C \wedge \forall\ s_i \in \mathcal{S}' \wedge (s_i, s_{i+1}) \in$ NSL:

$\mathcal{G}_{\text{Af}} : \mathcal{P}_C \to \mathbb{R}$

$\mathcal{G}_{\text{Af}}(\mathcal{P}) = \prod_{i=1}^{n-1} \mathcal{F}_{\text{Af}}(s_i, s_{i+1})$

This means that for a chain of network slices, the result for affinity is the product of values of this attribute for each of the links that belongs to the NSliceCh. □

Corollary 1. Affinity for a NSliceCh is achieved as a consequence of Property 7, since the NSliceCh is a subset of the CSG.

Trust (T):

It denotes the confidence to establish a business relation, enabled by the acknowledgment of the identity of the other party. Trust has an ordinal type of data, enabling to have levels of trust, for example, {trusted, not-trusted}, or equivalently, {1, 0}.

Property 9. Intuitively, the trust level of the destination NSlice has to be at least greater or equal to the trust level of the source NSlice.

We call $\mathcal{F}_{\mathrm{T}}$ the function that finds the trust for a link $(s_i, s_j) \in$ NSL.

$\mathcal{F}_{\mathrm{T}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$

$\forall s_i, s_j \in \mathcal{S}', \exists\ (a_{i_p}, v_{i_p}) \in \mathcal{A}(s_i) \wedge (a_{j_k}, v_{j_k}) \in \mathcal{A}(s_j) \mid a_{i_p} = a_{j_k} = \mathrm{T} \Rightarrow$

$$\mathcal{F}_{\mathrm{T}}(s_i, s_j) = \begin{cases} 1, \text{if } v_{ip} \geq v_{jk} \\ 0, \text{otherwise} \end{cases}$$

This means that if the trust of the links are at least the same, the function will have 1 as a result. □

Property 10. Trust level for a NSliceCh $\mathcal{P}$:

Let $\mathcal{C} = (\ \mathcal{S}'$, NSL). $\forall\ \mathcal{P} = \langle s_1, s_2, ..., s_n \rangle \in \mathcal{P}_C \wedge \forall\ s_i \in \mathcal{S}' \wedge (s_i, s_{i+1}) \in$ NSL:

$\mathcal{G}_{\mathrm{T}} : \mathcal{P}_C \to \mathbb{R}$

$\mathcal{G}_{\mathrm{T}}(\mathcal{P}) = \prod_{i=1}^{n-1} \mathcal{F}_{\mathrm{T}}(s_i, s_{i+1})$

This means that for a chain of network slices, the result for trust is the product of values of this attribute for each of the links that belongs to the NSliceCh. □

Corollary 2. The trust in a NSliceCh is obtained as an extension of the trust value in the links which embed it.

Security Level (SL):

It shows the rating of the NSlice in terms of security, for example its confidentiality, integrity or other criteria that can be measured. SL has an ordinal type of data, making possible to create, as its name implies, security levels to classify NSlice and manage their communication. For example, {high, medium, low}, or equivalently {3, 2, 1}. The intuition is that the SL of the destination NSlice has to be at least as high as the SL of origin NSlice

Property 11. We call $\mathcal{F}_{\mathrm{SL}}$ the function that finds the Security Level for a link $(s_i, s_j) \in$ NSL.

$\mathcal{F}_{\mathrm{SL}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$

$\forall s_i, s_j \in \mathcal{S}', \exists\ (a_{i_p}, v_{i_p}) \in \mathcal{A}(s_i) \wedge (a_{j_k}, v_{j_k}) \in \mathcal{A}(s_j) \mid a_{i_p} = a_{j_k} = \mathrm{SL} \Rightarrow$

$\mathcal{F}_{\mathrm{SL}}(s_i, s_j) = min(s_i, s_j)$

The outcome of this function is the minimum value of SL for the considered links. □

Property 12. Security Level for a NSliceCh $\mathcal{P}$:

Let $\mathcal{C} = (\ \mathcal{S}'$, NSL). $\forall\ \mathcal{P} = \langle s_1, s_2, ..., s_n \rangle \in \mathcal{P}_C \wedge \forall\ s_i \in \mathcal{S}' \wedge (s_i, s_{i+1}) \in$ NSL:

$\mathcal{G}_{\mathrm{SL}} : \mathcal{P}_C \to \mathbb{R}$

$\mathcal{G}_{\mathrm{SL}}(\mathcal{P}) = \min_{i=1}^{n-1} \mathcal{F}_{\mathrm{SL}}(s_i, s_{i+1})$

This means that for the NSliceCh the minimum value is used as a way to portray the lowest security level admitted on the path. □

3.5.2 Metrics

According to [89], a metric is "a standard of measurement that describes the conditions and the rules for performing a measurement of a property and for understanding the results of a measurement". A metric provides knowledge about an entity via its properties and the measured values obtained for that property. In our case, metrics are associated to links. For every link $(s_i, s_j) \in \text{NSL}$, it exists a metric vector m.

It is defined as: $m_{(s_i,s_j)} = \{(\text{Af}, \mathcal{F}_{\text{Af}_{(s_i,s_j)}}), (\text{T}, \mathcal{F}_{\text{T}_{(s_i,s_j)}}), (\text{SL}, \mathcal{F}_{\text{SL}_{(s_i,s_j)}})\}$.

3.5.3 Final remarks

After stating the attributes for network slices, the metrics and the functions to perform operations on them, the set of tools needed to validate a NSlice Chain is complete. The problem of the compliance with a security policy and the optimization of the process to find a suitable network slice chain is presented in the next section.

3.6 Policy validation for Network Slice Chains

The customer will specify the CS according to the intended use case scenario. This means that the customer will specify not only the service parameters, but also the required security constraints. On the CSP, the realization of these customer requirements depends on inter-slice communication. In other words, this communication should be regulated according to certain *rules* r_i that are grouped in a policy Π. The flow of data that will follow the chain of network slices must comply with the policy, in order to be authorized.

Specifically, rules are expressed as a vector $\langle$ Subject $\mathcal{SU}$, Object $\mathcal{O}$, Security Constraint Φ, Decision $\mathcal{DE}$ $\rangle$ and its components specify the conditions for communication. Subjects and objects were presented in Section 2.4.1; security constraints Φ in Section 2.4.3; and Permissions were addressed in Section 2.4.7. Likewise, the compliance operator $\cong$ presented in Section 2.5.4 will be used.

The policy defined by the customer is going to be used by the provider to select the most suitable network slice chain in its CSG, which complies with the service requirements and the security constraints. This will be addressed in the next Section. In order to ease

the understanding, the terms *path* and *network slice chain* refer to the same concept, unless stated otherwise.

3.6.1 Security constraint and optimization problem

It is necessary to verify that at least one NSliceCh, represented by $\mathcal{P}$, exists and complies with the metric in the policy. Φ corresponds to the constraints that must be respected, that is to say, that a path $\mathcal{P} \in \mathcal{P}_C$ in a CSG C matches the criteria if:

$\mathcal{G}_{\mathrm{Af}}(\mathcal{P}) \geq \Phi_{\mathrm{Af}} \wedge \mathcal{G}_{\mathrm{T}}(\mathcal{P}) \geq \Phi_{\mathrm{T}} \wedge \mathcal{G}_{\mathrm{SL}}(\mathcal{P}) \geq \Phi_{\mathrm{SL}}$

This means that not only the evaluation of each one of the attributes should be greater or equal than the ones specified by the constraints in Φ, but also that all those evaluations should agree. In order to perform the verification for $\mathcal{P}$, the CSP has to overcome two problems:

First, the CSP must be sure that the network slice chain offered to the customer respects the security constraints expressed by the customer in its policy. To do so, a service contract is established between them, in which the customer's required security policy is specified. The CSP has to assure that a network slice chain exists and complies with each one of the rules and, if it is not the case, notifies the customer to either, modify the security requirements, or kindly ask him to choose another CSP.

Second, the CSP must verify that the security resources used over the network slice chain are *as minimal as possible*. This can be interpreted as, for example, minimal operational cost, minimal calculation power or minimal traffic filtering capacity. This means that offering a network slice chain with security resources that are more powerful and of higher importance, will lead to a higher cost for the CSP. The consequence is a waste of resources, that could be instead be used on a service requirement for a more important customer that asks for a more important CS that provides more revenue to the CSP.

To optimize the use of its resources, the operator must minimize the error generated by the use of each link. Indeed, each link used in a path generates an error, which is the result of the difference in the security specifications between what the customer requires and the CSP offering. The greater the error, the more expensive the link used is for the CSP. To calculate the error introduced by the choice of the link, it is possible to use the Root Mean Squared Error (RMSE).

For a rule $\mathrm{r}_i = \langle\ s_1,\ s_n,\ \phi_{Af},\ \phi_T,\ \phi_{SL},\ \text{allow}\ \rangle$, the induced error for a NSL (s_i, s_j) is calculated as Formula (3.1):

RMSE: NSL $\rightarrow \mathbb{R}$

$$(s_i, s_j) \rightarrow \sqrt[2]{\frac{\left(\mathcal{F}_{\mathrm{Af}}(s_i,s_j)-\phi_{Af}\right)^2+(\mathcal{F}_{\mathrm{T}}(s_i,s_j)-\phi_T)^2+(\mathcal{F}_{\mathrm{SL}}(s_i,s_j)-\phi_{SL})^2}{3}} \quad (3.1)$$

From Formula (3.1) it can be deduced that the closer the error is to 0, the closer the link's specification will be to the customer's constraints and therefore to the service for which it is actually billed by the operator. On the contrary, the greater the error means that the link provides a higher level of security, it will be more expensive for the CSP and exceed the security expectations for the customer.

In function of Formula (3.1), it is possible to calculate the error of a path $p_i = \langle\ \mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_n\ \rangle \in \mathcal{P}$ that is a set of paths that have as source the NSlice s_1 and destination s_n with the formula 3.2:

ErrorPath: $\mathcal{P} \rightarrow \mathbb{R}$

$$\langle s_1, s_2, ..., s_n \rangle \rightarrow \sum_{i=1}^{n-1} \frac{\mathrm{RMSE}(s_i, s_{i+1})}{n-1} \quad (3.2)$$

What this means is that the error of the path is accumulated over all its constituting links. With this, we introduce a new problem called Security Constraint and Optimization Problem (SeCOP), wich is specified by Property 13.

Property 13. A CSP must offer for each rule $\mathrm{r}_i = \langle\ s_1, s_n, \phi_{Af}, \phi_T, \phi_{SL}, \text{allow}\ \rangle$ a path $p* = \langle\ \mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_n \rangle \in \mathcal{P}$, the set of paths between s_1 and s_n that:

— **(i)** Respect the constraints ϕ_{Af}, ϕ_T, ϕ_{SL}:
$\forall\ i \in [1, n-1]$, $\mathcal{F}_{\mathrm{Af}}(s_i, s_{i+1}) \geq \phi_{Af} \wedge \mathcal{F}_{\mathrm{T}}(s_i, s_{i+1}) \geq \phi_T \wedge$
$\mathcal{F}_{\mathrm{SL}}(s_i, s_{i+1}) \geq \phi_{SL}$

— **(ii)** Minimize the error over the path: $p* = min_{p \in \mathcal{P}}\ (Error_{Path}(p))$

□

In order to solve these problems, a new polynomial-time algorithm is proposed, as well as a method to find the optimal network slice chain, as is shown in the next Section.

3.7 Solving the challenges

The SeCOP problem presented in Section 3.6 via Property 13 has two parts: **(i)** compliance with security constraints; and **(ii)** optimization of resource utilization for the

path. Each of these parts can be solved by the algorithms presented in the following subsections. Figure 3.9 presents a summary of this process followed by an example, as a guiding aid to understand it better.

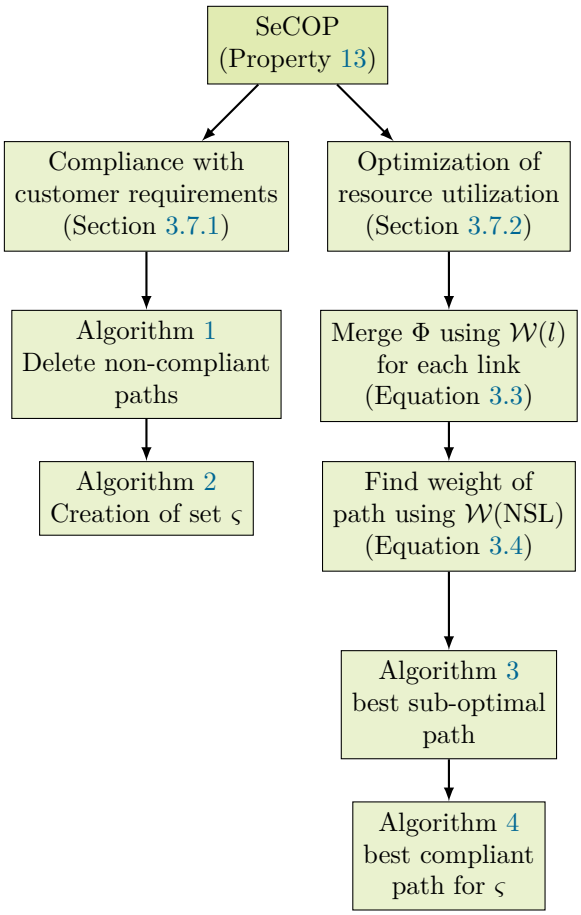

Figure 3.9 – Flow chart summarizing the process to solve the challenges.

3.7.1 Compliance with customer requirements

The first part of the SeCOP problem is to find all the paths that respect the security constraints: Affinity, Trust and Security Level. For a rule $\mathrm{r}_i = \langle\, s_1, s_n, \phi_{Af}, \phi_T, \phi_{SL}, \text{allow} \,\rangle$ a path p does not respect the constraints of one of its links i.e., that there is at least a link $(s_i, s_{i+1}) \in p$ **where**:

$$\mathcal{F}_{\mathrm{Af}}(s_i, s_{i+1}) \;<\; \phi_{Af} \;\vee\; \mathcal{F}_{\mathrm{T}}(s_i, s_{i+1}) \;<\; \phi_T \;\vee\; \mathcal{F}_{\mathrm{SL}}(s_i, s_{i+1}) < \phi_{SL}$$

Such a link each time it is used by a path will not guarantee compliance with the

constraints. It is of our interest to comply with the constraints, so it is necessary to represent those compliant paths separately. To represent the set of paths that satisfy the security constraints of a rule r_i, it is possible to transform the CSG C into a $\mathcal{C}_{ri}$ graph as stated by Property 14.

Property 14. A CSG $\mathcal{C}_{ri} = (\mathcal{S}, \text{NSL}_{ri}, \mathcal{F}_{\text{Af}}, \mathcal{F}_{\text{T}}, \mathcal{F}_{\text{SL}})$ with $\text{NSL}_{ri} \subset \text{NSL}$ which is a sub-graph of $\mathcal{C}$ where all the links guarantee the security constraints such that:

$$\text{NSL}_{ri} = \{l \mid l \in \text{NSL} \wedge \mathcal{F}_{\text{Af}}(l) \geq \phi_{Af} \wedge \mathcal{F}_{\text{T}}(l) \geq \phi_T \wedge \mathcal{F}_{\text{SL}}(l) \geq \phi_{SL}\} \qquad \square$$

With Property 14, the $\mathcal{C}_{ri}$ graph guarantees that any path between the source s_i and the destination s_n respects the constraints imposed by the client for this rule.

The process for transforming CSG C to $\mathcal{C}_{ri}$ for a r_i rule is given in Algorithm 1. The principle is to study each link of the C graph and to check that it respects all the security constraints. A link respecting these constraints is kept, otherwise it is deleted. The algorithm is polynomial because it has a complexity $O(n^2)$. Indeed, in the worst case, a directed graph has a link number $l = n \times (n-1)$. The $\mathcal{C}_{ri}$ graph respects the

Algorithm 1 Deleting non-compliant links

RemoveNonCompliantLink($\mathcal{C} = (\mathcal{S}, \text{NSL}, \mathcal{F}_{\text{Af}}, \mathcal{F}_{\text{T}}, \mathcal{F}_{\text{SL}})$, $r_i = \langle s, o, \phi_{Af}, \phi_T, \phi_{SL} \rangle$)
$\text{NSL}_{ri} \leftarrow null$
ForEach: $l \in \text{NSL}$
if $\mathcal{F}_{\text{Af}}(l) \geq \phi_{Af} \wedge \mathcal{F}_{\text{T}}(l) \geq \phi_T \wedge \mathcal{F}_{\text{SL}}(l) \geq \phi_{SL}$ **then**
  $\text{NSL}_{ri} = \text{NSL}_{ri} \cup l$
end if
End ForEach
Return $\mathcal{C}_{ri} = (\mathcal{S}, \text{NSL}_{ri}, \mathcal{F}_{\text{Af}}, \mathcal{F}_{\text{T}}, \mathcal{F}_{\text{SL}})$

constraints imposed by a r_i rule. The security policy Π contains n rules that have specific constraints. A $\mathcal{C}_{ri}$ graph must therefore be created for each r_i rule $\in \Pi$. We thus obtain a set $\varsigma = \{\mathcal{C}_{r1}, ..., \mathcal{C}_{rn}\}$ which includes all the graphs corresponding to the policy Π. Algorithm 2 provides the steps to create such a set ς. Once the graph transformations

Algorithm 2 Creation of set ς

CreateAllCompliantGraph(Π, $\mathcal{C} = (\mathcal{S}, \text{NSL}, \mathcal{F}_{\text{Af}}, \mathcal{F}_{\text{T}}, \mathcal{F}_{\text{SL}})$)
ForEach: $r_i \in \Pi$
$\varsigma = \varsigma \cup \text{RemoveNonComplianLink}(\mathcal{C}, r_i)$
End ForEach
Return ς

have been performed according to the security policy Π, the CSP can issue a warning to the customer about the rules that are not supported by the network. Certainly, a rule cannot be supported if the security constraints that it imposes are too demanding and no route that satisfies them can be found by the CSP. The set of Faulty Rules (FR) is given by Property 15.

Property 15. The set of rules $\mathrm{r}_i = \langle\ s_1,\ s_n,\ \phi_{Af},\ \phi_T,\ \phi_{SL},\ \text{allow}\ \rangle$ of a policy Π that the CSP cannot satisfy is:

$$\mathrm{FR} = \{r_i \mid r_i \in \Pi \wedge \nexists\ p = \langle\ \mathrm{s}_1,\ \mathrm{s}_2,\ \ldots\ ,\ \mathrm{s}_n \rangle \in \mathcal{C}_{ri}\}$$ □

This way, a set of graphs that contains the paths that comply with the policy is obtained solving the first part of the problem. To solve the second part of the SeCOP problem, it is necessary to optimize the use of security resources once it is known that a rule can be satisfied by a set of paths. To do so, a sub-optimal algorithm is proposed in the next Section to solve such a problem.

3.7.2 Optimization of resource utilization

Up to this point, we have a set ς of all the graphs containing the links that comply with the policy. In order to find a sub-optimal solution, the considered strategy is to merge the considered security attributes into a single numerical value using the expression in Formula (3.3):

$$\mathcal{W}(l) = \alpha\left(1 - \frac{\phi_{Af}}{\mathcal{F}_{\mathrm{Af}}(l)}\right) + \beta\left(1 - \frac{\phi_T}{\mathcal{F}_{\mathrm{T}}(l)}\right) + \gamma\left(1 - \frac{\phi_{SL}}{\mathcal{F}_{\mathrm{SL}}(l)}\right) \tag{3.3}$$

Where $\alpha + \beta + \gamma = 1$.

These coefficients are used to indicate the level of importance that an attribute has. This obeys the interest of the CSP according to its needs. Formula 3.3 establishes the per-attribute ratio between the constraint expressed by the customer and the value configured by the CSP. The definition of the scale for the attributes is open to be defined by the CSP.

The term *sub-optimal* is used because using the Formula (3.3) implies loosing information in the process of finding an aggregated metric to represent all the attributes.

The result of the evaluation of Formula (3.3) for each link is then used to find the

total weight of the path, using Formula 3.4:

$$\mathcal{W}(\langle s_1, s_2, ..., s_n \rangle) = \sum_{i=1}^{n-1} \mathcal{W}(s_i, s_{i+1}) \tag{3.4}$$

This way, a set of paths is obtained, all of them obeying the constraints expressed in the policy, merged on a single value. The set of paths are used as input for the Dijkstra algorithm presented as Algorithm 3, which will return the shortest-*distance* path from the initial set of paths. For our case, this *distance* is interpreted as the path with lowest weight found between the source and destination NSlice. This refers to the set of nodes whose links comply with the policy and have an aggregated metric $\mathcal{W}$, found using Formula 3.3. The final result indicates the best sub-optimal path that can be used to satisfy the

Algorithm 3 Finding best compliant path

```
function DIJKSTRA(Graph, Source, Destination)
    Q ← null
    for all each node v in Graph do
        dist[v] ← infinity
        prev[v] ← null
        add v to Q
        dist[source] ← 0
    end for
    while Q ≠ empty do
        u ← node in Q with min dist[u]
        remove u from Q
        for  each neighbor v of u do
            alt ← dist[u] + W(u, v)
            if alt < dist[v] then
                dist[v] ← alt
                prev[v] ← u
            end if
        end for
    end while
     return dist[], prev[]
end function
```

requirements from the customer.

In the case it is necessary to do this process for the set $\mathcal{S}$, the Dijkstra function in Algorithm 3 has to be called for each $\mathcal{C}_{ri}$ graph that belongs to $\mathcal{S}$. This is specified in Algorithm 4, which uses as input the policy Π (which contains the source and destination

NSlices in the rule) and the set ς.

Algorithm 4 Finding best compliant path for the whole set ς

```
function FindBestCompliantPaths(Π, ς)
    for all C_ri in ς do
        (src, dst) ← Π                    ▷ src and dst NSlices extracted from Π
        Dijkstra(C_ri, src, dst)          ▷ Algorithm 3 is called
    end for
     return dist[], prev[]
end function
```

3.7.3 Example

For a better understanding, a basic scenario is presented in Figure 3.10. The topology represents a CSP network that has a set of network slice instances, which conform a CSG, designed to provide a communication service. The layered structure of the network slices illustrates a real-world arrangement that CSP use in their networks, each layer providing a certain service to the next one.

The setup consists on 11 NSlices, being NSlice s_1 the source and s_5 the destination of the traffic. For this example, the CSP considers that the range of the attributes for Affinity (Af), Trust (T) and Security Level (SL) go from 1 up to 3. The unique rule states that the customer wants a NSliceCh that has $\phi_{Af} = \phi_T = \phi_{SL} = 2$.

On Figure 3.10, the numbers on each NSL show the metric for that NSL as configured by the CSP. For simplicity, that number applies equally to the three attributes ϕ_{Af}, ϕ_T and ϕ_{SL}. After executing Algorithm 1, the non-compliant links are removed. This way, only

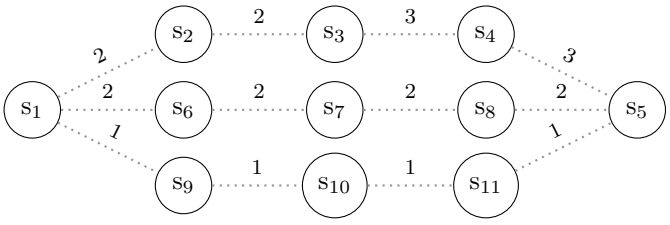

Figure 3.10 – Scenario to illustrate the proposed algorithm.

paths identified by $\langle s_1, s_2, s_3, s_4, s_5 \rangle$ and $\langle s_1, s_6, s_7, s_8, s_5 \rangle$ are kept. Execution of algorithm 2 is not necessary since there is only one rule in the policy.

The next step is to merge the metrics using Equations 3.3 and 3.4. It is assumed that the values for the parameters of this equations are fixed by the CSP to α =0.3, β =0.3, and γ =0.4.

For the links that have $\phi_{Af} = \phi_T = \phi_{SL} = 2$, an aggregated weight of 0.33 is obtained. For links that have $\phi_{Af} = \phi_T = \phi_{SL} = 3$, the weight value is 0.

After executing Algorithm 3 on the remaining NSliceCh with the weights that were found, and between source NSlice s_1 and destination NSlice s_5, NSliceCh 1 ($\langle s_1, s_2, s_3, s_4, s_5 \rangle$) has a weight of 0.33+0.33+0+0=0.66, while NSliceCh 2 ($\langle s_1, s_6, s_7, s_8, s_5 \rangle$) has a weight of 0.33+0.33+0.33+0.33=1.32.

From this, it is concluded that NSliceCh 1 is the best sub-optimal path that meets the security constraints from the customer.

But, a key observation is that this NSliceCh does not **exactly** comply with what the customer required. The customer required $\phi_{Af} = \phi_T = \phi_{SL} = 2$. The chosen NSliceCh uses higher specification of security resources, which are not being payed by the customer. In this case, NSliceCh 2 is the best path.

As this is a small example, the difference and the time spent to spot the difference between the best sub-optimal and best path is easy and done rapidly. But as the topology gets more complex, the calculation of the best path, the one that minimizes the error, that acknowledges exactly what the customer requires, takes a lot of time and it is not suitable to be used. This is shown via an implementation of this process in the next Section.

3.8 Implementation

In order to verify the proposed solution to the challenges a test-bed was set up. It was done via a program written in a Matlab R2020a, executed on a laptop with an Intel Core i5-7300U CPU at 2.6 GHz with 16 GB of memory. The test-bed consisted on a CSG with 100, 500, 1000 and 2000 slices, connected in a similar fashion as the example in Figure 3.10. For each of the network slice arrangements, an adjacency matrix was created with specified security attributes for Affinity, Trust and Security Level, which can be adjusted from 1 to 100. The policy stated that a path can be considered as compliant if the Affinity and Trust parameters are equal to 10, and if the Security Level is equal to 50. After the execution of the algorithms shown in Section 3.7, paths that meet the constraints are found. The analysis covers the execution time to find the best sub-optimal path and the calculation of RMSE and the Error Path. Then, with a table lookup it is

possible to know which network slice chain is closer to the security requirements from the customer. As the experiments showed, choosing a path with higher RMSE value is better in terms of security guarantees according to this test-bed scenario but, in practical terms, doing so leads to a waste of resources on the CSP part, because it goes beyond of what was desired by the customer.

3.8.1 Execution time

Figure 3.11 shows the results of our tests to find the time it takes to execute the best path i.e., the path p* that minimizes the error (see Property 13). It is shown with a blue line. As expected, as the number of NSlices increases, the time to calculate the best

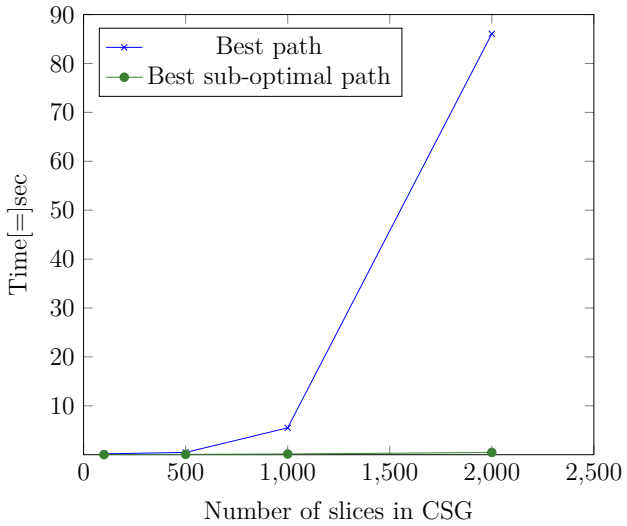

Figure 3.11 – Time to find a communication service path for the considered scenarios.

sub-optimal path increases. This calculation is represented with a green line. For the CSG containing 100 network slices, 0.006 sec were needed; for 500, 0.066 sec; for 1000, 0.148 sec; and for 2000 network slices, 0.46 sec. This covered the execution of Algorithms 1, 2 and 3. This time increase also applies to the process of finding the best path, because to obtain this information it is needed to scan the adjacency matrix for all paths from source to destination. In the case of the CSG consisting of 100 network slices, 0.181 sec were needed; for 500, 0.46 sec; for 1000, 5.49 sec; and for 2000 NSlices, 86 sec were needed. Evidently, finding the best path is not a suitable solution for real-time applications. Regarding the number of valid paths, the trend is to grow as the number of NSlices grows, because more connections are available for the traffic to flow. This is depicted by a blue line in

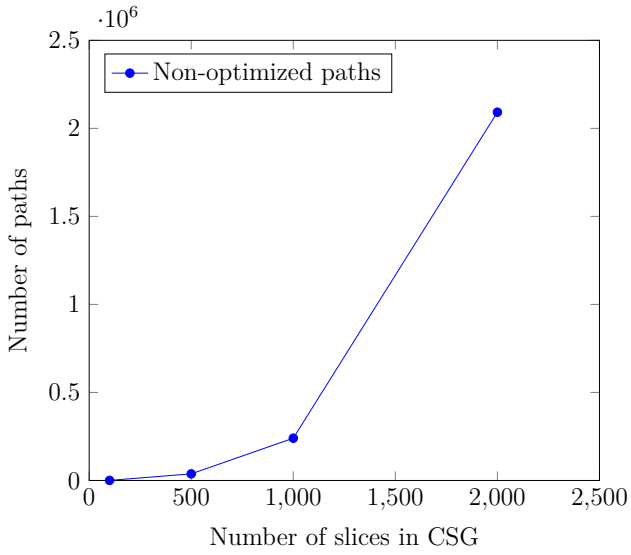

Figure 3.12 – Number of non-optimized of paths for the considered scenarios.

Figure 3.12 where, for the test-bed with 2000 slices, almost 2.1 million paths were found. This is an interesting outcome: there are so may available paths that it is important for the CSP to optimize resources according to security attributes. The conventional approach to optimize a path according to, for example, performance, is no longer sufficient.

Analyzing further the execution times, in the case of the setup with 100 NSlices, the needed time to determine the best sub-optimal time was 0.006 sec. For the case of 2000 NSlices, the needed time was 0.46 sec. A 20 times increment in the number of slices, had as consequence of almost 77-times increase in the processing time. This insight of the scaling in the amount of network slices with respect to the delay for the policy evaluation is interesting, because it must be considered in the operator's time-budget analysis. These results depend on the particular characteristics of the implementation, but it is useful in order to have an idea of the growth as the test-bed increases in size of network slices, number of considered attributes, their scale of values and complexity of the policy evaluation.

3.8.2 Comparison between best and sub-optimal communication service path

Since no path complied exactly with the customer's requirements as stated in the policy, an estimation of the deviation of the parameters of the compliant paths from the

desired ones can be performed. This error estimation is important in order to provide to the CSP more tools to take decisions: it reflects how close is a network slice chain to satisfy the exact customer requirements regarding security. Figure 3.13 shows a cloud of

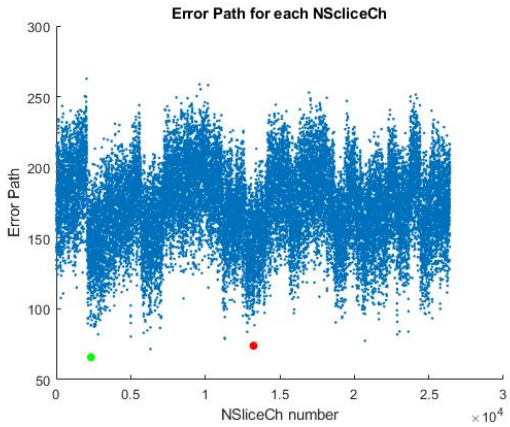

Figure 3.13 – Error Path calculation for 500 network slices.

measurements of the Error Path value for each one of the found network slice chain for the test-bed of 500 network slices. On it, there are two points of interest: one green point in the lower left side (the best path p*), and a red point (the sub-optimal path found with our algorithm) higher than the green point at the center of the Figure. The lowest the measurement, the lowest the error measurement is, meaning that path is really close to meet exactly the security requirements expressed by the customer. If there is no network slice chain that match the requirements (like in our experiment), the one represented by the green point is the closest one. The other network slice chain marked in red is the best one according to Dijkstra, the one with best weight according to the linear function used to merge the parameters, but its requirements regarding security could be exceeding the resources needed to meet the needs of the customer. A similar analysis can be made for the test-bed containing 2000 network slices, shown in Figure 3.14. The difference between the two Error Paths for the interest points is 6.76. This measurement is valuable as a tool to provide more information about the alignment of the security settings of a slice with respect to a policy.

3.8.3 Final Remarks

The implemented test-bed reflects a real-world scenario that will become common for a CSP. Customers will specify a communication service along with security attributes

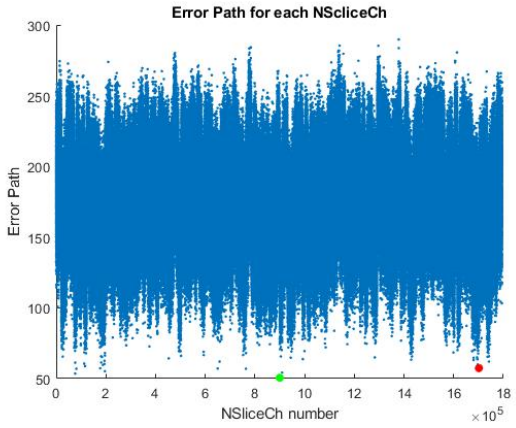

Figure 3.14 – Error Path calculation for 2000 network slices.

according to their needs.

It is important for the CSP not to optimize resources based only on traditional attributes as performance or bandwidth. In fact, security can be seen as a new service for the CSP, who will be able to charge its customers for it. This constitutes new sources of revenue for the CSP.

Regarding the implementation of security, it should not become an additional operational and expensive load for the CSP. In order to avoid this situation, the CSP should find out how to provide a secured service to a customer with the lowest cost.

This can be achieved by using our proposed algorithm to find a sub-optimal path, because calculating the best path takes a lot of time. Once the best path is computed it will be used instead of the sub-optimal path. However, such a strategy can only work when the client's security policy varies little over time. Likewise, when there are important topological changes due to the dynamic nature of the network, the sub-optimal path remains the only usable solution.

Another important message is that the error path calculation gives an idea of how dispersed are the sub-optimal paths with respect to the security parameters stated by the customer. Some paths can be better suited for the task because they exceed the security expectations from the customer, at expense of resource detriment for the operator. This tool is helpful to identify the network slice chain that either fit the security constraint exactly or is the closest to it.

3.9 Discussion

The utilization of network slices as a mechanism to provide communication services to customers and tenants will become commonplace, as technology becomes mature and adoption of enabling technologies such as NFV and SDN increases. Since the nature of a network slice is conceived as a unit specially assembled for a certain use case, the creation of rich end-to-end communication services necessarily involves the communication between several network slices, constituting a network slice chain.

Besides the customization of network functions and services, communication services must comply with security requirements. These requirements are described as attributes, which characterize a particular security constraint that must be preserved along the chain of network slices that will deliver the service.

These elements are used in the communication model that **(1)** assures that there is a network slice chain connecting the required network slices that complies with the constraints expressed in the policy; **(2)** assures that beyond compliance, the offered path has minimal security-related resource utilization for the CSP.

First, a network slice model was proposed, demonstrating the concept of network slice chain. The resulting inter-slice communication model specifies key elements to manage the security when connecting network slices in a consecutive fashion. This model is extensible for application in any service and for the inclusion of other security attributes, so security requirements can be expressed more richly. It complies with any access control model, ensuring a straightforward implementation.

Then, a new polynomial time algorithm was proposed, which takes into account the security attributes for the service as established by the policy, and via a linear function merges the values into a single one. With this value, a sub-optimal network slice chain was found using Dijkstra's algorithm.

This algorithm was tested using a network of 100, 500, 1000 and 2000 network slices connected in a layered fashion, just as a real-life deployment in a CSP network. Besides the service specification, each network slice had specific security attributes for affinity, trust and security level. The obtained delays during testing are promising in order to validate chains of network slices that comply with policy in real-time. To do so, a strategy that can be adopted is to pre-compute all valid paths, to then choose the best one.

Last, by calculating the RMSE value between the security metrics of all found network slice chains against the reference values in the policy asked by the customer, it was possible

to find that there are paths that exceed the security expectations from the customers. Choosing those paths, even though it is better from the security point of view, are at expense of waste of resources on the CSP side. This is due to the fact that the function used to merge the attributes lead to lose information and, in consequence, misleading the decision making of the most suitable network slice chains for the customer.

Chapter 4

Metrics to assess the isolation of Network Slices

4.1 Introduction

As presented previously in this document, network slicing has the power to host several services over the same infrastructure, enabled by an intelligent resource sharing. Some communication services components would require to be alone in the infrastructure, while others could be hosted together. This leads to the concept of isolation, which is inherent to network slicing and refers to the degree of resource sharing that could be tolerated by the industry partner [90]. For the CSP, this is a key feature, because its objective is to maximize the usage of its infrastructure to host as many tenants as possible. This way, to have high revenue as the infrastructure is utilized at its highest capacity. However, some caution has to be taken regarding infrastructure sharing: **(i)** assure coexistence of the services without degrading performance; **(ii)** restrict interaction between tenants (e.g. traffic); and **(iii)** guarantee fairness in performance, ensuring that the customer gets according to what he is paying for. It is necessary to keep in mind that improper isolation leads to security problems, due to the shared infrastructure [90].

In order to implement isolation in a proper way, the CSP has to overcome important challenges, such as: **(i)** defining isolation according to the situation of the service; **(ii)** verifying the concerned layer: service, data plane, control plane, or management; **(iii)** identifying the involved isolation categories that have to do with security, which could refer to performance, traffic, bandwidth, storage, CPU, among others; and **(iv)** establishing properly its specification (how is it constructed) and its measurement with proper units and thresholds.

The challenge is that there is no consensus about the parameters needed to assess the isolation of network slices and that there are no methods to measure the isolation of a communication service instantiated via network slices. In consequence, there is no metric

to make the bench-marking of **(i)** the isolation for a communication service and **(ii)** the isolation level of the network slices deployed in the infrastructure. This information is useful to help the CSP to instantiate the components of the CS in a way that respect internal policies regarding isolation and keep the SLA with the customer.

To solve this, our proposition is to provide: **(i)** the list of parameters to measure the secure isolation of a network slice; **(ii)** the process to calculate the isolation level of a network slice, which then can be used as a way for comparison with another network slice, and **(iii)** the process to assess the isolation of the communication services offered by the CSP via network slices.

This chapter is organized as follows: Section 4.2 describes the state of the art on the subject under discussion; Section 4.3 proposes an architecture to address the challenges; and Section 4.4 provides the considered elements to assess the secure isolation of network slices. Section 4.5 presents the considerations regarding mapping of the considered elements of the data structure to calculate the metric. Section 4.6 illustrates the process to perform the calculation of the metric, to demonstrate then in Section 4.7 its implementation in a test bed scenario. Final remarks on the subject are presented in Section 4.8 followed by a final discussion in Section 4.9.

4.2 Recent work

The isolation specification is a complex topic since it can be achieved at different levels, according to the resource at which we would like to provide such assurance [33]. Some key parameters and areas that could be configured accordingly to provide isolation are:

— Provide separate hardware for the CS.
— Segment the traffic using Virtual Local Area Network (VLAN).
— Dedicate bandwidth using Traffic Engineering (TE) techniques and protocols like Multi-Protocol Label Switching (MPLS) [91].
— Assign dedicated addressing space for the CS.
— Use virtualization techniques such as virtual machines, hypervisors, or containers [92].
— Perform storage segmentation using Storage Area Networks (SAN).
— Use application-level utilities or Virtual Private Network (VPN) applications.

Other authors prefer to consider the isolation from the point of view of the datacenter and the cloud computing paradigm. In [93], authors propose isolation as a major challenge

for a data center operator and list several technologies and protocols to provide such isolation at host and core levels. Nonetheless, most of the work is about performance isolation [94] since at this moment, the usage of generic "white boxes" does not provide assurance of a networking service performance compared to when it is implemented on a dedicated equipment.

One of the risks of the implementation of isolation from the CSP point of view is to experience under-utilization of resources, especially on the RAN [95]. Into network slicing literature, there are few publications that address this subject and the efforts for its quantification have been minimal [39].

Measuring the isolation is very important, because it helps to assess SLA compliance, measure the security level, resiliency and high availability of a CS. It also provides valuable information to:

— Allocate slices according to the requirements in isolation.
— Manage the security and isolation of network slices.
— Decide on life-cycle management actions on network slices.
— Survey the state of the network.

Even if recent research works state which parameters to use to specify it, the works that focus on its measurement are minimal.

These ideas lead to think about a base architecture to seek to manage the security isolation of a network architecture. ETSI in [96] proposes a Security Manager (SM) for the NFV-MANO reference architecture, which aims to support security monitoring and management. Even though ETSI does not address concrete metrics or addresses a concrete security problem or isolation scheme, its virtualized architecture can be enhanced to solve the challenges about the isolation assessment.

Authors in [94] propose three metrics to quantify the level of performance isolation focused on the Software as a Service (SaaS) cloud model. They cover the QoS impact, workload variability (increase or decrease, according to the customer behaviour) and an integral metric. They focus on virtualized systems and do not cover an integral approach regarding physical considerations on isolation. Authors in [40] address the problem of the optimal allocation of a slice in a 5G core network, by considering the physical isolation requirements between the components of the slice. Specifically, by placing the VNF optimally to provide intra-slice isolation. But their approach does not consider the CSP point of view regarding the isolation of all instantiated services or on providing a metric to benchmark isolation parameters.

On patent [97] they propose a method for managing network slices in a communication network. They are focused on the shared NF part. They propose the use of a flag "service isolation". For them, it is about grouping the slices that can share NF and exclude the ones that do not want their NF to be shared, but no specification for the quantification of the isolation is given. Authors of patent [98] go a step further and include isolation requirements as part of deployment strategies for network slices, but they do not specify how to use these parameters to enforce isolation or even measure it. In patent [99] they mention about isolation levels (hard and soft isolation) and provide the idea to compare the isolation requirement with the isolation capability of the network. Depending whether the requirements are higher or lower than the ones the network can provide, the slice is allocated. They do not provide the logic or process to find out the isolation level or to perform the comparison.

To our understanding, CSP have methods to **apply** isolation, usually for:

— Hardware: using redundant hardware equipment.
— Traffic: applying segmentation by configuring VLANs or creating broadcast domains.
— Bandwidth: applying QoS or traffic engineering using MPLS-TE.
— Virtualization: using hypervisors to share resources between virtual hosts by using scheduling and sharing algorithms.
— Storage: creating logical unit numbers (LUN) for a storage area network (SAN).
— Application: using Muti-Tenant Applications.

These mechanisms to apply isolation are well understood, but there is no method to **measure** it. Intuitively, it could be done by gathering all the required parameters and compute them together to get a single value that reflects the isolation ranking of the network slices. The process developed to solve the challenges covers:

— Find the required parameters that constitute the metric for isolation of network slices,
— Combine those parameters according to a concrete methodology to quantify the metric for the isolation,
— Present an implementation to illustrate the application of the proposed method.

This process has its foundations in an architecture over which services are deployed using physical and virtual resources. The architecture acts as a guide to visualize where the parameters for isolation are located. This is presented in the next Section.

4.3 Architecture

ETSI in [45] establishes the foundation of the NFV architecture and its management. The architecture in Figure 4.1 shows its different layers of interest.

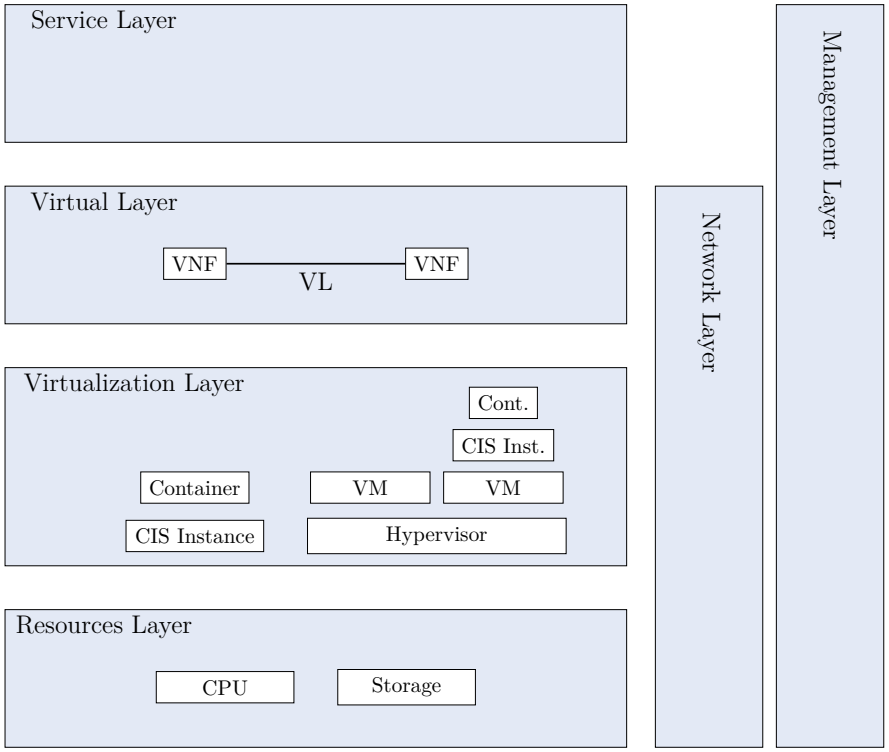

Figure 4.1 – Architecture that aims to guide the process to find metrics for isolation.

— Service Layer: It is the one where all the services are realized, so they can be consumed by customers. This realization could be done, for example, via network slice instances.
— Virtual layer: It refers to the "space" where VNF are instantiated offering a service. It leverages on resources that are provided by the virtualization technologies offered by the lower layer.
— Virtualization layer: It contains virtualization technologies over which VNF are instantiated. Technologies cover hypervisors and Container Infrastructure Service Instance (CISI), which can host Virtual Machines (VM) and containers, respectively.
— Resources layer: It refers to the infrastructure, which contains the physical resources, such as CPU, storage and network.

— Management layer: Provides orchestration and monitoring capabilities to the aforementioned layers. It has the monitoring tools to harvest the information that is required. To do so, they query the Management Information Base of the components of the network in order to get information such as bandwidth, alarms, packets loss, among others. This information is valuable for the CSP in order to take actions.
— Network layer: Provides required connectivity as well as to address multi-tenancy. It is transversal to the virtual, virtualization and resources layers.

The network layer is an abstraction of the network as a tool for connectivity that goes beyond the capabilities that are provided by the hypervisors and CISI. Although the hypervisor abstracts the Network Interface Card (NIC) so all the guests can communicate with each other via an internal virtual switch, the need for connectivity between guests residing over different hypervisors is not addressed. This way, the network layer offers enhanced capabilities to connect the network slices that will provide the services.

Having the understanding of the layers that compose the proposed enhanced architecture, next Section shows the data that is needed to obtain the metric for isolation.

4.4 Data Structure

In order to quantify the isolation, it is necessary to organize the information available, leveraging on the way the different layers interact. Each layer (L_x) has attributes (A_i), which refer to a feature or property of an entity. In our case, the features that are important are the ones related to isolation, which must have a way to be assessed via metrics. The challenge is the difficulty to come up with all the metrics. To answer to this need, we created *Metric Categories* ($MetCat_j$) that help to provide a set of groups of *Metrics* (Met_k) that are related to an attribute. This logical construction is seen in Figure 4.2. The way to quantify the isolation is determined by using a function whose parameters are the Metric Categories for each attribute. The values used for each metric (or even the values used for a metric category) are assigned by the CSP according to its expertise, to the use case scenario and to the range of possible configurations. As candidate metrics are added as needed, a richer set of parameters is assembled to quantify the isolation.

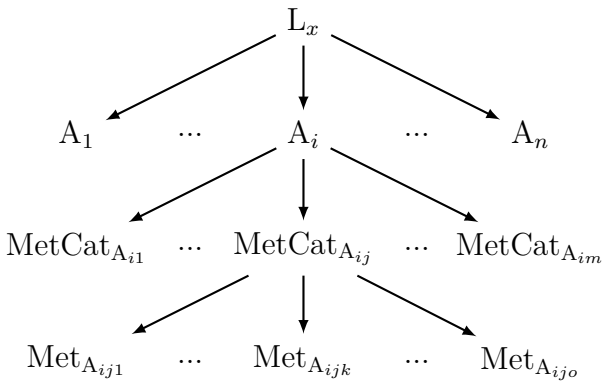

Figure 4.2 – Structure of the composition of attributes, metric categories and metrics for a generic layer.

4.4.1 Attributes (A)

As said earlier, attributes refer to the features or properties of a layer, which can be extended to the entities that are within the layer. The name of the considered attributes are: function, dissociation, security, traffic and specification. The description of these attributes is presented as follows:

- Function: Refers to the purpose of the layer as a whole or an individual entity, the role it performs, the service it provides or the type of traffic it processes.
- Dissociation: Gives an indication about the affinity of the entity to share a resource with another entity, according to its service or partitioning scheme. Also gives information about the performance that must be satisfied for the operation of the entity, the manufacturer of the equipment and the equipment family as well as the placement of it.
- Security: Provides information about the overall security objectives that are inherent to the entity.
- Traffic: Describes the characteristics of the flow that is processed by the entity.
- Specification: Gives information about the nature and characteristics of the entity used to provide a service within a layer.

4.4.2 Metric Categories (MetCat)

Finding all metrics to measure and quantify isolation is very difficult. That is why the approach is to find categories to group metrics for each attribute. For each category it is necessary to find concrete and representative metrics for it. An important remark here is

that an attribute can be used in several layers. For example, access control can be used as a security attribute of the service layer and the virtualization layer. In a similar fashion, a Metric Category can be used in several attributes. For example, performance, which can be applied to traffic or hardware characteristics such as load or response-time guarantees for dissociation. A non-exhaustive list of metric categories (grouped by attribute) is shown as follows.

Function

— Container/VM function: has metrics to evaluate the functionality of a container or a VM.
— Life expectancy: measures the expected operational duration of an entity.
— Role: denotes the function of an entity.
— Service type: helps to differentiate the service offering according to its characteristics.
— Traffic type: helps to differentiate the traffic according to its nature or to whom it belongs (user plane, control plane or management)

Dissociation

— Affinity: denotes the criteria by which a policy can enforce or deny the sharing of network, link and storage resources, as well as sharing the same physical location.
— Location: denotes where an entity is. It can indicate, for example, the data-center it is placed, or the server a network function is instantiated.
— Segmentation: denotes the differentiation of the technologies used to confine features of the entities.

Security

— Access Control: contains metrics that measure the robustness of access control model.
— Accountability: contains metrics that assess the requirement for actions of an entity to be traced uniquely to that entity.
— Availability: groups the metrics used for protection against: **(i)** deletion of data; **(ii)** unauthorized use of resources. Likewise, availability refers to the usage of the resources by an authorized entity, being these resources performing its function for

an interval of time.

— Confidentiality: metrics that permit to assess: **(i)** protection from intentional or accidental attempts to perform unauthorized data reads; **(ii)** data or information is not made available or disclosed to unauthorized entities.

— Integrity: metrics to demonstrate that data has not been altered at rest or in transit. Also, for a system, reflecting the logical correctness and reliability of that entity.

— Privacy: contain metrics to analyze how information is handled. Examine and evaluate protections and alternative processes for handling information to mitigate potential privacy risks. Actions related to invasion of privacy: **(i)** appropriation, **(ii)** intrusion, **(iii)** public disclosure of private information.

— Reliability: measures of the ability of an item to perform a required function under stated conditions for a stated period of time.

— Trust: contains parameters that help to assess the reliance on the integrity, business, and compliance of an entity.

— Life cycle management: metrics to evaluate the functioning state of the entity.

Traffic

— Bandwidth: contains the result values of the measurement of the data transfer rate, bit rate or throughput.

— Protocol: metrics to compare the different types of implementations of a service.

— Security Level: denotes characteristics of traffic according to its confidentiality, integrity, availability and privacy requirements.

Specification

— Compliance: metrics that measure agreement with regulations and standards.

Miscellaneous

— Performance: contains metrics to measure "how many, how fast" type of requirements.

— Provider: refers to ways for identifying different types of providers of service or equipment.

4.4.3 Metric (Met)

According to NIST in [89], a metric is a standard of measurement that describes the conditions and the rules for performing a measurement of a property and for understanding the results of a measurement. A metric provides knowledge about a property through both its definition (e.g., expression, unit, and rules) and the values resulting from the measurement of the property. Metrics can be used as desired, in order to assess the level of isolation that is required.

4.4.4 Summary tables

Summary Tables 4.1, 4.2, 4.3, 4.4, 4.5 and 4.6 show an example of the layers, attributes, metric categories and selected metrics.

Table 4.1 – Examples of the considered metrics for the service layer.

Layer	**Attribute**	**metric category**	**Metrics**
Service	Function	Service type	Slice/Service Type (SST): eMBB, URLLC, and mMTC.
	Dissociation	Affinity	Slice/Service Type (SST), VLAN.
	Security	Access Control	Criteria to allow or deny interactions between slices: their role, domain, or the owner of the service.
	Traffic	Security Level	Function of confidentiality, integrity, availability and privacy.
		Protocol	Type of protocol (secured, not secured, level of security, etc)

Table 4.2 – Examples of the considered metrics for the virtual layer.

Layer	Attribute	metric category	Metrics
Virtual	Function	Service type	Slice/Service Type (SST): eMBB, URLLC, and mMTC.
		Role	Type of VNF: customer, service or governance roles.
		Traffic type	Whether traffic is in the control plane, user plane, or management plane.
	Dissociation	Affinity	Slice/Service Type (SST), VLAN.
		Performance	Response time, vCPU load, throughput, bit rate, and latency.
	Security	Trust	Attestation, VNF certification, TCB, behavior, identity and date.
		Access Control	Usernames, roles, attributes, technology to exchange credentials, for identification and authentication.
		Privacy	Encryption methods: AES, and the number of bits.
		Availability	Calculated via a mathematical formula, reliability schemes: active-active; active-backup; geographical redundancy.
		Integrity	Checksums, data validation.
		Confidentiality	Sensitivity level, encryption.
		Reliability	Mean time before failures (MTBF).
		Accountability	Role of the entity, non-repudiation schemes.
	Traffic	Performance	Delay, jitter, packet loss.
		Security Level	Function of confidentiality, integrity, availability and privacy
		Protocol	Type of protocol (secured, not secured, level of security, etc)
		Bandwidth	Throughput.

Table 4.3 – Examples of the considered metrics for the virtualization layer.

Layer	Attribute	Metric category	Metrics
Virtualization	Function	Container/VM function	Type and specific function of the entity performing virtualization.
		Life expectancy	Instantiation and termination time.
	Dissociation	Affinity	Service Type: whether the intended task agrees with the type of service provided.
		Performance	Response time, Sharing algorithm, Density, CPU load.
	Security	Trust	Software authenticity, TCB, behavior, identity, date.
		Integrity	checksums, data validation, Trusted Platform Module (TPM) and virtual TPM (vTPM).
		Availability	Realibility schemes.
		Access control	Usernames, roles, attributes, technology to exchange credentials, for identification and authentication.
		Accountability	Role of the entity, non-repudiation schemes.
		Confidentiality	Sensitivity level, isolation schemes, scheduling strategies, time multiplexing, space multiplexing.
		Reliability	Mean time before failures (MTBF).
	Specification	Provider	Virtualization technology, type, brand, speed.
	Traffic	Performance	Delay, jitter, packet loss.
		Bandwidth	Throughput.

Table 4.4 – Examples of the considered metrics for the network layer.

Layer	Attribute	Metric category	Metrics
Network	Dissociation	Segmentation	Technology, protocol: IP, ATM, Frame-Relay, PPP, BGP, OSPF.
	Security	Confidentiality	Encryption scheme using public-key and symmetic-key.
		Privacy	Encryption methods: RSA, AES, TripleDES
	Specification	Provider	Brands, models, switches, routers, Virtualization technology, type, brand, speed of the SDN controller, or the virtual switches.
	Traffic	Traffic performance	Delay, jitter, packet loss.
		Bandwidth	Throughput.

Table 4.5 – Examples of the considered metrics for the hardware layer.

Layer	Attribute	Metric category	Metrics
Resources	Dissociation	Provider	Brand of the servers, switches, and other equipment.
		Segmentation	Technology, brand, protocol, access control technique, zoning.
		Location	City, neighborhood, power provider diversity, transport network provider diversity.
	Security	Availability	Realibility schemes.
		Confidentiality	Isolation schemes, scheduling strategies, time multiplexing, space multiplexing, encryption scheme.
		Integrity	Checksums, data validation, TPM.
		Privacy	Encryption methods.
		Reliability	Mean time before failures (MTBF).
		Trust	Attestation, TCB, TPM, behavior, identity, date.
	Specification	Provider	Brands, models, type of HW, architecture, technology.
	Traffic	Performance	Delay, jitter, packet loss.
		Bandwidth	Throughput.

Table 4.6 – Examples of the considered metrics for the management layer.

Layer	Attribute	Metric category	Metrics
Management	Security	Accountability	Role of the entity, non-repudiation schemes.
		Trust	Trust level, trust type. It uses roles, domains, attestation, certificates.
		Privacy	Encryption methods.
		LCM	Current state of the entity, permissions to perform LCM actions.
	Specification	Provider	Brands, model, architecture, programming language of the implementation.
		Compliance	SDO API compliance.

4.5 Considerations about mapping

As presented in Section 4.4, there is a hierarchy relation between the Layers, Attributes, Metric Categories and the Metrics. The process to establish this parent-child relationship is not an easy task. This Section provides guidelines about this problem.

4.5.1 Mapping of metrics to metric categories

The association of metrics to the metric category can be made by direct mapping, as suggested previously. Nonetheless, since there could be a lot of metrics according to the use case, an automatic approach is needed. Machine Learning techniques can help to find distinguishable qualities on metrics and group them accordingly into a metric category. Moreover, besides the method to perform the mapping, a function is needed to have a normalization of the values of the metric into a [0,1] interval. The normalization function in stated in Property 16 as follows:

Property 16. We call $\mathcal{F}_{\text{MetCat}}$ the function that normalizes the value of the metric $met \in$ MetCat.

$\mathcal{F}_{\text{MetCat}} : \text{met}^k \rightarrow [0, 1]$

The outcome of this function is the corresponding mapping and normalization of the selected options for the metric *met* into a [0,1] interval. □

Definition 7. There exists at least a function for each MetCat, which combines the metrics and normalizes them into the [0,1] interval. □

An example of the rationale used by the CSP for the election and normalization of those values is presented in Table 4.7. For each metric, possible values or ranges of values are conceived, as well as the outcome of the normalization process. These values can come after technical or market evaluation according to the CSP expertise, can be stated according to the specifications of the infrastructure, the characteristics of the metrics and the way the services are implemented and configured in the network. There is no global silver-bullet for the value-assignment and the outcome of the normalization function: each CSP is autonomous in its approach to do this.

For instance, for throughput, the CSP can start this analysis from the physical layer. Most of data centers use fiber cables that are either 10 Gbps, 40 Gbps or 100 Gbps. If a customer needs 10 Gbps, the CSP will assign a single fiber for that service of that single customer. This means that entire physical resource is for that customer only. If a customer

needs 1 Gbps or 2 Gbps, it is likely the cable that the customer will use will be shared with other customer of a comparable demanded capacity. This sharing of physical media contributes to lower the isolation level for that service.

Table 4.7 – Rationale for the value election for the considered metrics.

Metric Category	**Metric**	**Possible values**	$\mathcal{F}_{\mathbf{MetCat}}$	**Rationale**
Trust	Behavior	Low; Medium; High	0.1; 0.5; 0.9	Trust on the entity translates in higher security perception
Bandwidth	Throughput	T <1 Gbps; 1≤ T <10 Gbps; T ≥ 10 Gbps	0.3; 0.7; 0.9	Higher throughput requires using dedicated physical links and ports in equipment. This means higher isolation.
Confidentiality	SecLevel	Low; Medium; High	0.1; 0.5; 0.9	Higher security level translates into more sensitive payload level. Requires higher isolation.
Affinity	SST	eMBB; mMTC; uRLL	0.2; 0.5; 0.8	Low latency requires higher performance, in consequence, high isolation.
Privacy	Encryption	None; SSL 128 bits; SSL 256 bits	0 ; 0.5; 0.9	Higher encryption provides higher privacy rating and more security.

In order to illustrate analytically the normalization, let us take for example the attribute *security*, which has several metric categories, such as confidentiality, trust or privacy. Specifically for *confidentiality*, we considered two metrics: sensitivity level and encryption (abbreviated as met_1 and met_2 in Equation 4.1). The CSP can establish the sensitivity level using a classical scale of high, medium and low according to the characteristics of the resource. Likewise, the encryption can be ranked by type of algorithm and its number of bits, for example, Advanced Encryption Standard (AES) with 128-bit, 192-bit and 256-bit. Intuitively, higher the sensitivity and number of used bits in the algorithm reflects a higher confidentiality. The function according to Property 16 is:

$$\mathcal{F}_{\text{Confidentiality}}(\text{met}_1, \text{met}_2) = \frac{max(value(\text{met}_1), value(\text{met}_2))}{\text{Met}_{max}} \tag{4.1}$$

With $value(\text{met}_1), value(\text{met}_2) \in [1, \text{Met}_{max}]$

The function in the example uses the maximal function to combine the values of the metrics, but it is up to the CSP to define this operation according to its business strategy and objectives. Thanks to Equation 4.1, a unique normalized value can be found when a

metric category contains several metrics.

4.5.2 Mapping of attributes to layers

Not all attributes relate to all the layers or the entities inside the layers. The following figures show the envisioned mapping of the parameters to the architecture. There are two cases that are considered: **(i)** some attributes apply to the whole layer, that is, all its internal entities; and **(ii)** several attributes are specific to some of the entities inside the layer. Either way, the mapping analyzes the required attributes that are necessary to characterize each one of the layers and its internal entities.

Figure 4.3 depicts the mapping for the first case. For instance, the Service Layer

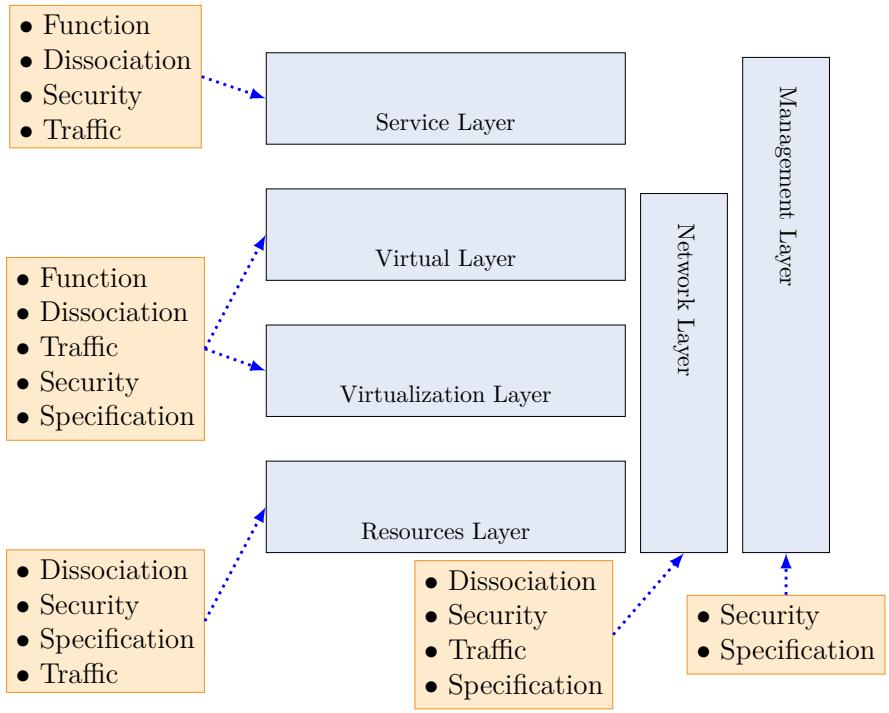

Figure 4.3 – Matching between each attribute to a whole layer.

requires a specification of the *function* that the service has, the *Dissociation* has to share lower level resources, the specification of its *security* requirements and the type of *traffic* it will handle. For the rest of the layers, an important attribute is *specification*, which helps to achieve isolation by ensuring a degree of diversity, by specifying resources that may perform the same function but can be provided by a PNF or VNF supplied by different manufacturers, brands or software companies. This way, a bug or vulnerability

in a platform will not compromise the whole service as there are entities with the same function that are provided by another manufacturer, which are not vulnerable to the same problem.

In similar fashion, Figure 4.4 portrays the mapping for the second case. In particular,

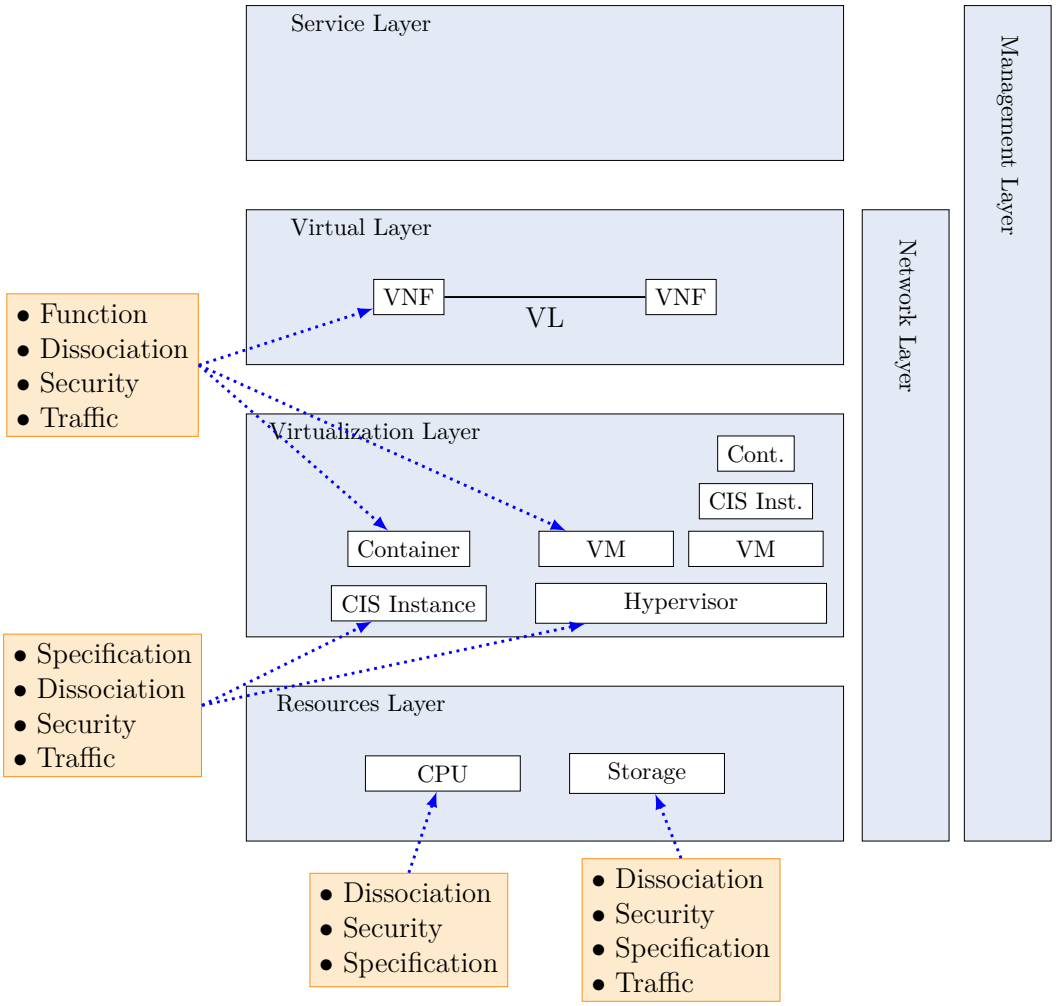

Figure 4.4 – Matching between each attribute to a specific entity within a layer.

for entities that have virtualization capabilities, such as the CSI Instance and the hypervisors, the *Dissociation* attribute plays an important role, in order to assess the level of affinity of the instantiated containers or VM under their scope. Regarding lower layers, the *Traffic* attribute is important to have control of the level of the segmentation of the traffic and its characteristics.

4.6 Calculation of metric for isolation

The idea is to find a way to perform the quantification of the isolation, considering the attributes of each layer. Layers, attributes, and metric categories can have weights, which can be applied in two ways according to the calculation model (shown in Sections 4.6.1 and 4.6.2). The weight denotes the degree of importance in the contribution to isolation enforcement. The approach is flexible so that the CSP can choose to only use a subset of all the attributes or metrics, just by zeroing the weight that is not needed. The convention to denote the weight of each of these parameters is: the weight for a layer ($\mathcal{W}_{L_x}$), the weight of an attribute ($\mathcal{W}_{A_i}$), and the weight of a metric category ($\mathcal{W}_{\mathrm{MetCat}_j}$). For simplicity, by now, the value denoted by each attribute depicts the final isolation value for the metric categories considered for that attribute. For an easy visualization, attributes and layers are abbreviated as shown in Tables 4.8 and 4.9.

Table 4.8 – Layers: abbreviations and assigned weight according to the CSP point of view.

Layer	$\mathbf{L}_x$	$\mathcal{W}_{L_x}$
Hardware	L_1	0.3
Virtualization	L_2	0.2
Virtual	L_3	0.2
Service	L_4	0.05
Management	L_5	0.05
Network	L_6	0.2

Table 4.9 – Attributes: abbreviations and assigned weights according to the CSP point of view.

Attribute	$\mathbf{A}_i$	$\mathcal{W}_{A_i}$
Function	A_1	0.1
Dissociation	A_2	0.2
Security	A_3	0.3
Traffic	A_4	0.2
Specification	A_5	0.2

The calculation process proposes the use of linear functions (the use of non-linear functions is also valid, but out of the scope of this work). In this method for the calculation of isolation, the idea is to vary where the weights are applied: either on the attributes or on the metric categories. The decision to choose the option to apply rests with the CSP. This, depending on its business purpose, its domain of expertise and desire to enforce a special property for the services. Either way, the CSP could also consider the mix of the

two options, but this approach is out of the scope for this thesis. These two options are explained as follows.

4.6.1 Option 1: weight on attributes

In this case, the weights are assigned to the attributes of the layer, as seen in Figure 4.5. Since each attribute A_i is composed of m metric categories, all of used metric categories

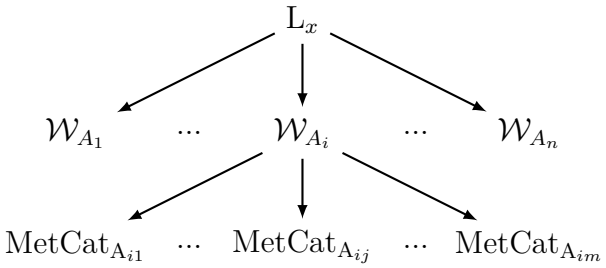

Figure 4.5 – Weight on attributes.

will be sharing the same weight $\mathcal{W}_{A_i}/m$. In consequence, all the metric categories have the same level of importance to the attribute.

In order to find the isolation for a given A_i, Equation 4.2 can be used:

$$I_{\mathrm{A}_i} = \mathcal{W}_{A_i} \times \sum_{j=1}^{m_{\mathrm{A}_i}} \mathrm{MetCat}_{\mathrm{A}_{ij}} \tag{4.2}$$

Where m_{A_i} refers to the m number of metric categories under the scope of A_i.

Then, the total isolation for a layer L_x can be found using Equation 4.3 as follows:

$$I_{\mathrm{L}_x} = \sum_{i=1}^{n} I_{\mathrm{A}_i} \tag{4.3}$$

It is important to keep in mind that the sum of weights of all the n attributes of the layer must be equal to 1, as per Equation 4.4:

$$\sum_{i=1}^{n} \mathcal{W}_{A_i} = 1 \tag{4.4}$$

4.6.2 Option 2: weight on Metric Categories

In this case, weights are specific to a metric category. This means that some metric categories can have more importance than others. As it is presented in Figure 4.6, for the attribute A_i, each of the metric categories has its own weight. An interesting question here

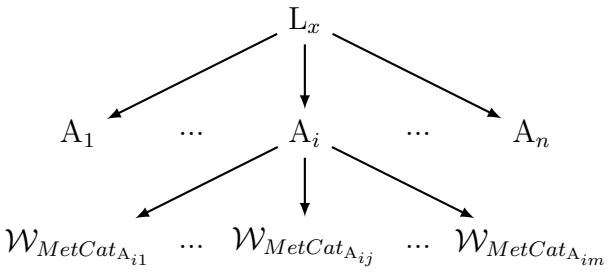

Figure 4.6 – Weight on metric categories.

is how to aggregate the values of the inner metrics that belong to different metric category. This is answered by the use of the function described in Property 16 and Definition 7. The CSP has the power to decide the nature of the function according to its interest. For instance, for the *trust* metric category, the CSP can use the minimal value; for *privacy*, the maximal value. After the decision from the CSP, the isolation for A_i can be found using Equation 4.5:

$$I_{\mathrm{A}_i} = \sum_{j=1}^{m_{\mathrm{A}_i}} \mathcal{W}_{MetCat_{\mathrm{A}_{ij}}} \times MetCat_{\mathrm{A}_{ij}} \tag{4.5}$$

Then, for the total isolation for a layer L_x Equation 4.3 can be used again.

It is important to keep in mind that the sum of weights of all the m metric categories of the attribute must be equal to 1, as per Equation 4.6:

$$\sum_{j=1}^{m} \mathcal{W}_{\mathrm{MetCat}_j} = 1 \tag{4.6}$$

where m is the number of metric categories.

4.6.3 Total calculation

After choosing one of the two options to apply the weight, we can continue to calculate the total isolation value. For it, we consider that each layer L_x may have a level of

importance, that is, a weight $\mathcal{W}_{L_x}$, for the contribution to isolation. The value for this weight obeys the internal CSP criteria. For its assignment, the intuition is that:

— Independent elements between services demonstrate higher isolation level.

— Independent elements in lower layers of the architecture is good for nothing if there are shared elements on top of the considered layer.

Adhering to option 1 and using Equation 4.3, the total isolation I_{Total} for all the layers of a service is expressed as shown in Equation 4.7:

$$I_{Total} = \sum_{x=1}^{6} \mathcal{W}_{L_x} * I_{\mathrm{L}_x} \tag{4.7}$$

Being x the variable that represents the number of levels. A practical example of this process is presented in the next Section.

4.7 Implementation

To demonstrate the concept, two service requests are received from the customers, considering a metric for each metric category (MetCat) in each attribute (A) for each layer (L).

4.7.1 Topology

The topology in Figure 4.7 shows the mapping performed by the CSP for the service requests from the customers. The mapping strategy is out of the scope of this thesis, so we trust the mapping reflects the best interests for the service offered to the customers. The purpose is to illustrate the pairing between the customer request for the service against the available types of VNF that are already instantiated to provide a service. On the figure, the mapping for Customer 1 is represented by a dashed green line, for Customer 2 with a dashed yellow line, and the mapping from the virtual layer (L_3) to the virtualization layer (L_2) with a dashed blue line. Customer 1 specifies that it desires an eMBB service, with 0.5 Gbps of traffic capacity, a medium trust environment with an encryption scheme that uses SSL 128 bits. Customer 2 also asks for an eMBB service, but with 1 Gbps of traffic, a high level of trust and encrypted using SSL 256 bits.

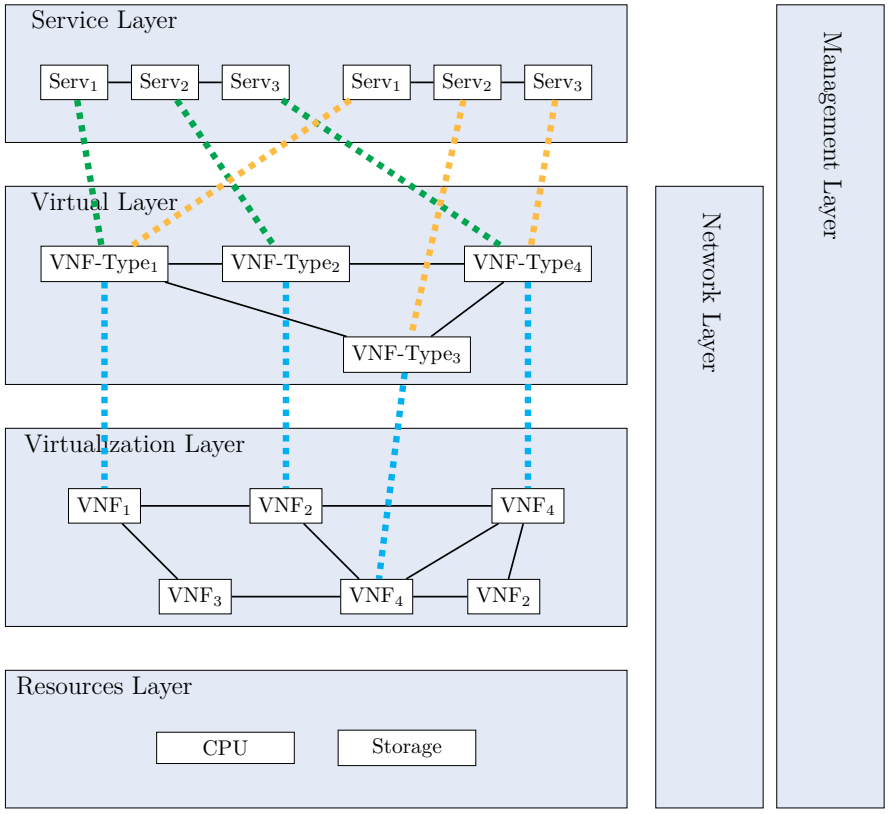

Figure 4.7 – Example using the proposed architecture.

These requirements are summarized in Table 4.10 where the cells with a dash (-) denote that the layer is not involved as a requirement for the customer as part of the service specification.

Table 4.10 – Reference table for the selected configuration options for each customer. Each selected option corresponds to a metric category, which belongs to an attribute at a specific layer.

$\mathbf{L}_x$	$\mathcal{W}_{L_x}$	$\mathbf{A}_i$	$\mathcal{W}_{A_i}$	**MetCat**	**Selected Metric**	**Customer 1**		**Customer 2**	
						Selected Option	$\mathcal{F}_{\mathbf{MetCat}}$	**Selected Option**	$\mathcal{F}_{\mathbf{MetCat}}$
L_1	0.3	-	-	-	-	-	0	-	0
L_2	0.2	A_3	0.3	Confidentiality	SecLevel	Medium	0.5	High	0.9
L_3	0.2	A_4	0.2	BW	Throughput	0.5 Gbps	0.3	1 Gbps	0.7
		A_3	0.3	Trust	Bahavior	Medium	0.5	High	0.9
L_4	0.05	A_1	0.1	Service Type	SST	eMBB	0.2	eMBB	0.2
L_5	0.05	-	-	-	-	-	0	-	0
L_6	0.2	A_3	0.3	Privacy	Encryption	SSL 128	0.5	SSL 256	0.9

4.7.2 Infrastructure capabilities

The capabilities of the virtual entities and physical infrastructure are known by the CSP beforehand. Table 4.11 provides an example of the attributes, considered metric categories and the selected option for the corresponding metric value, assigned by the provider. The metric value shall be normalized later in order to be useful to find the total isolation for a service. Each layer has a contribution as a whole to the isolation. This contribution can be seen from two points of view: **(i)** the intuition that if the isolation mechanisms are in a lower layer (towards the physical equipment), those mechanisms are more effective in their contribution to isolation. This is reflected in the weight value assigned to each layer (as shown in Table 4.8); and **(ii)** weights can be assigned according to the use case and the interest of the provider. For example, a CSP focused on financial services would be more interested in providing a higher weight to the upper layers, starting from the virtualization layer.

As shown in the topology, the Network Layer (L_6) is "transverse" to the Hardware Layer (L_1), Virtualization layer (L_2) and Virtual Layer (L_3). This Network Layer is important for connectivity of the entities, specially in multi-tenant scenarios. The Management Layer (L_5) is assigned to a lower weight value because it could be considered as providing lower isolation in comparison to L_6. The Services Layer (L_4) is on the top, following the

Table 4.11 – Description of the infrastructure capabilities.

$\mathbf{L}_x$	**Name**	**Entity**	$\mathbf{A}_i$	**Name**	**MetCat**	**Metric**	**Selected Option**
L_3	Virtual layer	VNF-Type$_1$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	High
			A_4	Traffic	Bandwidth	Throughput	1 Gbps
		VNF-Type$_2$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	Medium
			A_4	Traffic	Bandwidth	Throughput	0.5 Gbps
		VNF-Type$_3$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	High
			A_4	Traffic	Bandwidth	Throughput	1 Gbps
		VNF-Type$_4$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	Medium
			A_4	Traffic	Bandwidth	Throughput	1 Gbps
L_2	Virtua-lization Layer	VNF$_1$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	High
			A_3	Security	Confidentiality	SecLevel	High
			A_4	Traffic	Bandwidth	Throughput	1 Gbps
		VNF$_2$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	Medium
			A_3	Security	Confidentiality	SecLevel	Medium
			A_4	Traffic	Bandwidth	Throughput	0.5 Gbps
		VNF$_3$	A_2	Dissociation	Affinity	SST	IoT
			A_3	Security	Trust	Behavior	Low
			A_3	Security	Confidentiality	SecLevel	Medium
			A_4	Traffic	Bandwidth	Throughput	200 Mbps
		VNF$_4$	A_2	Dissociation	Affinity	SST	eMBB
			A_3	Security	Trust	Behavior	Medium
			A_3	Security	Confidentiality	SecLevel	High
			A_4	Traffic	Bandwidth	Throughput	1 Gbps
		VNF$_5$	A_2	Dissociation	Affinity	SST	IoT
			A_3	Security	Trust	Behavior	High
			A_3	Security	Confidentiality	SecLevel	Medium
			A_4	Traffic	Bandwidth	Throughput	500 Mbps
L_6	Network Layer	-	A_3	Security	Privacy	Encryption	SSL

same line of reasoning. Evidently, these weights can change according to the business and point of view of the provider.

4.7.3 Scales, possible values and value assignment to metric categories

According to the possible options that can be selected for the metric, and subject to the technical evaluation of how well it contributes to the isolation level of a service, a value for each MetCat_j is found by the CSP thanks to the normalization function specified in Property 16. This value varies from 0 to 1, numbers that provide a sense of the level of isolation. For our example, **0** means very low isolation level and **1** means very high isolation. The practical meaning of this concept depends of the context and business purpose of the customer. Towards the physical layer, it can express the level of independence in resource utilization and how private the used resources are. Private physical resources not only perform better but are less prone to disturbance produced by customers that share the underlying resources. At the virtual layer, isolation can also refer to the number of competing processes that are running on a virtual CPU inside a VNF. The more exclusive are those virtual CPU, processes can run with better performance, without delays and less security risks. This isolation scale gives a sense of that level of independence and operation with the less disturbance as possible. It is useful in order for the CSP to take action: trigger corrective measures to guarantee the isolation requests for the customers. The value for each MetCat_j is found for each selected metric option for the service for each customer, as shown in Table 4.10.

4.7.4 Calculation

The weight on the attributes is assigned by the CSP according to its focus and business case. Table 4.10 summarizes the information concerning Customer 1 and Customer 2. The reader may observe that the sum of the values of the columns that contain the weight is not equal to one. The reason is because the customer's request only contains a subset of attributes and metric categories that the CSP supports in their service portfolio. If a customer does not specify a parameter, it is zeroed for that specific customer. Being said, the CSP must assure that the sum of the weight values for the whole attributes and metric categories that are for their interest is equal to one. The weight values must not change since it would jeopardize the fair calculation of the weight for all the set of customers of

the CSP.

Another important remark on Table 4.10 is that on the customer specifications only one metric was chosen. In consequence, the value of the selected option is mapped directly to the corresponding metric category. If the customer had specified several metrics for a single metric category, as stated in Section 4.6.2, their aggregation would obey the CSP strategy, such as using the minimum or maximum value, or an average.

Applying the mathematical expressions presented in Section 4.6, and concretely Equations 4.2, 4.3 and 4.7, it is possible to combine the weights of each of the layers of interest, with the weight of the attribute and the value for each of the considered metrics. Applied for each metric for each customer:

$I_{Total_{Cust1}} = 0.2 \times (0.3 \times 0.5) + 0.2 \times (0.2 \times 0.3) + 0.2 \times (0.3 \times 0.5) + 0.05 \times (0.1 \times 0.2) + 0.2 \times (0.3 \times 0.5) = 0.103$.

$I_{Total_{Cust2}} = 0.2 \times (0.3 \times 0.9) + 0.2 \times (0.2 \times 0.7) + 0.2 \times (0.3 \times 0.9) + 0.05 \times (0.1 \times 0.2) + 0.2 \times (0.3 \times 0.9) = 0.191$.

In the service-to-service comparison, the service for Customer 2 has better isolation qualities that are demonstrated with a better isolation index compared to the service of Customer 1. In practical terms, this means that the service for Customer 2 has a higher level of independence compared to the one for Customer 1. Consequences of this higher isolation are, for example, better performance and less threats to privacy, due the better security specifications.

Our approach feeds the isolation index with metrics linked to security. This could lead to think that a service that uses strong encryption algorithms, hard confidentiality and privacy schemes would be better isolated. But remember that this is just one part of the scheme. If the used underlying resources are shared, this fact lowers the isolation index. The other way around is also possible: a service that uses dedicated hardware and dedicated links can be implemented with low level security parameters, that can render its isolation level to a bad ranking.

Changing the parameters of a service changes its isolation index. This is an added value of having a quantitative measure of isolation: the CSP can perform again the comparison with the other indexes of the other services, in order to know if there are constraints that are not respected, and make changes accordingly. For example, move a service from one server to another one, in order to respect the the customer's service specification regarding isolation.

4.8 Final remarks

Using metrics is really important in order to support decision-making for the CSP as well as for:

- Helping to select services supported over network slicing among a set of possible providers,
- Defining and enforcing service level agreements,
- Monitoring services provided via network slices,
- Accounting and auditing the contracted services.

Specific to isolation, the presented methodology helps the customer to properly specify this requirement to its provider. With this information, the CSP can take action in several ways, for example:

- Judge whether the demand can be fulfilled according to the isolation requirements,
- Analyze how isolated are two services, and if required, migrate a service to other infrastructure in order to increase its isolation index,
- Perform planning of the deployment according to a desired isolation index proposed by the customer,
- Instantiation of services according to the affinity or anti-affinity of services inside the infrastructure,
- Provides a sense of where a service can be instantiated, via the mapping of the requirements into the deployed infrastructure characteristics.

The developed methodology is complete enough to provide tools for the stakeholders of the 5G ecosystem to quantify isolation and use this outcome to make accurate decisions in order to increment its level and modify the provided services accordingly.

4.9 Discussion

Isolation is one of the principal qualities inherent to network slices. Isolation is important in order to assure security of the services and ensure independent resources used by the services that use the same infrastructure. Communication service providers have the mechanisms to provide isolation at different levels of the infrastructure. It can be via deploying independent hardware, or segmenting traffic, by creating different broadcast domains, applying traffic engineering, among other strategies. Their challenge is to have a quantitative way to measure the isolation, in order to assess the isolation state of the

offered network slices, and perform decisions in order to enhance this index when needed.

To solve this challenge, Metric Categories were created, which seek to state a framework of important qualities that are necessary to provide isolation. These qualities depend on the layers over which the service is provided having as reference the ETSI NFV architecture. This is a flexible approach since getting all metrics is an impossible task. Then, a system of weights is implemented, to reflect how the different attributes contribute to the isolation. All of them are added as a linear function in order to find its total contribution to isolation, that is, its isolation index value.

This quantitative value helps the communication service provider to take intelligent decisions about the isolation state of its network slices. Actions can be, for example, to perform migration of services to a different server in order to increase its isolation index or segment traffic or reroute it to avoid bandwidth competition.

The isolation index constitutes a valuable key tool for the communication service provider to enhance the service offer via network slices.

Chapter 5

Conclusion and Perspectives

5.1 Conclusion

This document presented a pathway to solve the proposed challenges about network slicing security and its security management, specifically about network slicing isolation in a 5G System.

First, a theoretical background was constructed, seeking to clarify the definition of what a network slice is. Based on definitions provided by standard definition organizations, a global description of a network slice was created. On that, a research was developed in order to unveil its challenges regarding security. To overcome the weaknesses, a security architecture was proposed, containing essential components to control not only the instantiation and deployment of network slices, but also the monitoring of their security and mechanisms to mitigate risks and solve security incidents. A great effort was made to cover the future scenarios for utilization, and having the entities that help to keep the services and infrastructure secured. However, aiming to solve all the security challenges is unrealistic, so a choice was made to focus on the inter-slice isolation and the management of the security, using the proposed architecture as a starting point for development of the solution.

The suggested architecture is general enough to cover and incorporate the propositions from other companies and SDO, being extensible to adopt monitoring entities that are useful to assess the security state of the network slices.

Second, a new secure access control model called RDAC was proposed, which incorporates the best qualities from the traditional access control models. Crafting this access control model was a great challenge, because it was necessary to understand the interactions between elements in the 5G system to then find the issues that current access control models have when porting them into the 5G service based architecture. Traditional access control models were initially conceived for information system environments and the way security constraints are specified are limited to that use case. More advanced models are

stricter in the classification and hierarchies of the information and flows, but using them would make the interactions in 5G more difficult and expensive in time and processing, rendering the system unusable for our purpose. To reach an understanding, key properties from those access control models were chosen, being those properties fundamental for the functioning of the 5G system. Those properties reflect the principal qualities that are used to assess and authorize interaction in the target architecture.

As was stated in our published paper [7], by including key concepts as role, domain and security constraints, the proposed access control model assures a tight control of the interactions between entities in the 5G system. The model leverages on the specification of hierarchies for domains and classification of roles, in order to provide high granularity in the construction of rules in the policy to govern actions. The model is extensible so new properties can be incorporated and can be applied to other types of architectures and applications.

As a third step, a zoom-out of the architecture was made, going from considering interactions *inside* the 5GS network slice, into a *broader* point of view in order to consider the inter-slice communications. This is important with the purpose of conceiving enriched communication services that can be offered by stitching together several network slices from different types. The communication service providers benefit from this initiative, since there are other use cases that can be addressed by interconnecting single-purpose network slices generating more business opportunities and other sources of income. In addition, it enables the provider to reuse already-deployed network slices, permitting to save resources.

However, network slice interconnection must obey security constraints, the ones specified by customers compared to the ones that are configured in the network from the provider. Specifically, the model seeks to evaluate how well secured is the provided service, providing information whether the operator is using an excess of security resources over a service that does not deserve so much usage of them. This problem, which we called the Security Constraint and Optimization Problem (SeCOP) was solved via a novel polynomial-time algorithm. As it was presented in our publication [9], the proposed model provides assurance that not only a path that interconnects all the required network slices exists, but also that all candidate paths comply with the security constraints expressed in the policy. In addition, the algorithm finds the path that uses less security resources.

This insight is useful for the operator to manage optimally its security resources and to use more important security network functions to services or customers that are willing

to pay for them. This contribution was tested with an implementation that proved that for a 2000 network slice set-up, the experimented delays to find the network slice chains that complied with policy are short enough to be used in real-time scenarios. However, delays have to be analyzed according a specific use case scenario.

Last, the problem of quantification of the isolation of network slices was addressed. This is a management problem, that seeks to provide awareness to the service provider and the customer about how well isolated a network slice is compared to other network slices in the infrastructure. A methodology to use metrics inherent to isolation was created, seeking for simplicity of the calculation but giving the option for a more complex approach according to the needs of the involved stakeholders. This methodology was formalized to file a patent that is under study at this moment. This isolation index is an input for the security manager, which can help to take action to improve the isolation of the service, contributing to increase its security. The service provider can use this isolation index value to evaluate the trade-off between **(i)** a chain of slices that complies strictly with the policy (but takes a lot of time to calculate); **(ii)** with another network slice chain that is fast to calculate (but consumes a lot of security resources); and **(iii)** with another slice chain that is very well isolated (but does not exactly comply with the policy). These pieces of information can be used to feed a learning model in order to be analyzed and reach the best decision.

This work covered part of the challenges that are inherent to the deployment of network slices. This leads to think about future work on these topics.

5.2 Perspectives

Telecommunications are a field that is continuously evolving through time, due to the new use cases that need to be addressed. Security, as an inherent quality that has to be conceived along with the new services, is no exception and must keep the pace. There is research to be done on topics such as enable the security awareness of network slices, enhance the protection of the network slices from the subscriber misbehavior (intentional or unintentional), provide adaptive security to network slices via online traffic steering. Nonetheless, next steps that can be considered towards the enhancement of the security of 5G network slices are:

Impact analysis of securing network slices: When dealing with security measures, most of us would be in favor of deploying all the available security resources in

order to be sure nothing will damage our network services. It is necessary to take a step back and reflect about the consequences of the measures used to do so. Leaving aside the communication service, deploying security countermeasures has a direct impact on **(i)** the cost of the service (implies using more resources, and somebody has to pay for them); **(ii)** the network overhead (traffic has to be steered, it has to traverse other VNFs, additional encapsulations are used); **(iii)** the QoS management (additional delay, diminished throughput and performance, poor user experience); and **(iv)** the service management (service becomes more complex, additional network functions and slices have to be appended or inserted in the service path). The challenge is to analyze the consequences of actions under the economic and service level scopes, in order to choose the most suitable defense for the service. Our work provides first steps into the resolution of this challenge, because it permits to estimate whether the chain of slices complies exactly with what the customer requires. As outcome from this evaluation, it gives an idea of the excess of resources that are used to secure the service with the instantiated slices. Via an stricter and multidimensional assessment (considering other QoS metrics) a more accurate cost evaluation can be made. After assessing the impact of a mechanism to secure a network slice, it is necessary to deploy the mechanism into production by using a life cycle management scheme.

Include security vision into the LCM of network slices: In [19] 3GPP established the process to instantiate a network slice. This life-cycle management process is seen under the point of view of the service. Nonetheless, the confidentiality, integrity and availability have to be preserved for any slice instantiated by this process. For it, the inclusion of the aforementioned security features has to be assured from *preparation and design phase*, to be later enforced in the instantiation and configuration phases. In fact, the LCM process must include a verification that the service request meets minimal security specifications in order to be accepted for *instantiation and activation*, that is, its deployment. This is important because it allows the network slice to begin working with all the security measures that are needed, diminishing the risks of security breaches. While in *run-time* phase, that is, in service, the network slice instance as a "living entity" must have an awareness of the security events that are happening. Since the network slice cannot embed all types of functionalities, an external security manager is needed, in order to keep track of these situations. They will have a closed-loop relationship that will help to update policies and enforce them in real time, as the events happen. The challenge is to develop the characteristics of this feedback loop: update frequency, trigger mechanisms,

monitoring filtering, dynamic access control model and context awareness. A possible way to solve this challenge is to embrace the use of the Zero-Trust Network paradigm [100], that considers these features. Likewise, machine learning mechanisms can help to filter events and render more accurate the process to trigger corrections. As a related challenge, there are concerns about the authorized entities that can trigger LCM actions on network slices. API may use secure authentication to validate the entity that issues the request, but authorization schemes are needed in order to narrow down the actions that can be performed, the type of network slice that must be used and security strategy. These request will be received by an orchestrator, that will configure and instantiate the necessary resources in the infrastructure.

Placement of a security VNF: Among the features that are provided by SDN and NFV we find the ability to deploy a security VNF, that is a VSF. This feature can be used to support the enhancement of the security of a network service that is provided via network slicing. Two initial problems that need to be addressed are **(i)** selecting the *location* where a VSF must be deployed in order to fulfill a security response task, and **(ii)** choosing the right *specification* of the VSF capabilities to fulfill its job.

Besides placement and specification of the VSF, there must be a careful consideration on the *deployment* strategy [101]. A CSP can choose to use **(i)** online deployment (the VSF is placed as the request arrives, leading to optimization problems) or **(ii)** offline deployment (queuing several requests and deploying them in batch, achieving optimal placement at expenses of a delay in service). The CSP must also analyze the consequences of adding VSF to a network slice in the sense of the increase of the service footprint and attack surface. The multi-domain and multi-tenant scenario must be analyzed too, since communication services that involve several stakeholders will become commonplace. In this scenario, one of the challenges is on the mutual agreement on the attributes, metrics and location parameters that are needed to specify the VSF instantiation.

The placement choice depends on the type of problem to solve, the available resources in the infrastructure and the possible consequences on the service delivery as the security VNF is part of the service. To help to answer to this challenge, it is necessary to have awareness of the underlying resources, the service characteristics and the layer at which the security function is needed. Our work can be considered as part of the solution of this problem, in order to include a network slice that contains security capabilities. This way, the VSF that is needed can be included in the service. This involves the modification of the network slice chain that holds the service in order to include the network slice that

provides security services.

These challenges demonstrate that the pathway into securing network slices is not over. More specialized solutions are needed in order to address current and future problems. However, each problem solved paves the way into a more secure service and enhancing trust for customers.

Chapter 6

Publications from the thesis

L. Suárez, D. Espes, P. Le Parc, F. Cuppens, P. Bertin, and C.-T. Phan, **"Enhancing network slice security via Artificial Intelligence: challenges and solutions,"** in Conférence CESAR 2018, Rennes, France, Nov. 2018.

L. Suárez, D. Espes, P. Le Parc, and F. Cuppens, **"Defining a Communication Service Management Function for 5G Network Slices,"** in 2019 European Conference on Networks and Communications (EuCNC): Network Softwarisation (NET) (EuCNC 2019 - NET), Valencia, Spain, Jun. 2019.

L. Suárez, D. Espes, F. Cuppens, P. Bertin, C.-T. Phan, and P. Le Parc, **"On an access control model enhancement for the 5G system,"** in 2020 european conference on networks and communications (EuCNC): Posters (EuCNC 2020), Dubrovnik, Croatia, Jun. 2020.

L. Suárez, D. Espes, F. Cuppens, P. Bertin, C.-T. Phan, and P. Le Parc, **"Managing Secure Inter-slice Communication in 5G Network Slice Chains,"** in Data and Applications Security and Privacy XXXIV, Springer, Conference on Data and Applications Security and Privacy (DBSec 2020), Regensburg, Germany, Jun. 2020.

L. Suárez, D. Espes, F. Cuppens, P. Bertin, C.-T. Phan, and P. Le Parc, **"Formalization of a security access control model for the 5G system,"** in 2020 11th International Conference on Network of the Future (NoF) (NoF 2020), Bordeaux, France, Oct 2020.

Filed patent number PCT/FR2020/050817: **"Method For Quantifying The Isolation Of A Network Slice For 5G Networks"**

BIBLIOGRAPHY

[1] Cisco, *Annual Internet Report*, 2020 (cit. on p. 13).

[2] Akamai, *Global State of the Internet Security and DDoS Attack Reports*, 2019 (cit. on p. 14).

[3] NGMN, *5G White Paper*, 2015 (cit. on pp. 15, 26).

[4] Jason Elliot and Sameer Sharma, *Dynamic End-to-End Network Slicing Unlocks 5G Possibilities*, 2016 (cit. on p. 15).

[5] Luis Suárez et al., « Enhancing Network Slice Security via Artificial Intelligence: Challenges and Solutions », *in*: *Conférence C&ESAR 2018*, Rennes, France, Nov. 2018 (cit. on p. 19).

[6] Luis Suárez et al., « Defining a Communication Service Management Function for 5G Network Slices », *in*: *2019 European Conference on Networks and Communications (EuCNC): Network Softwarisation (NET) (EuCNC2019 - NET)*, June 2019, DOI: 10.1109/EuCNC.2019.8802010 (cit. on p. 19).

[7] Luis Suárez et al., « On an Access Control Model Enhancement for the 5G System », *in*: *2020 European Conference on Networks and Communications (EuCNC): Posters (EuCNC2020 - Posters)*, Dubrovnik, Croatia, June 2020 (cit. on pp. 19, 138).

[8] Luis Suárez et al., « Formalization of a Security Access Control Model for the 5G System », *in*: *2020 11th International Conference on Network of the Future (NoF) (NoF 2020)*, University of Bordeaux, France, Oct. 2020 (cit. on p. 19).

[9] Luis Suárez et al., « Managing Secure Inter-Slice Communication in 5G Network Slice Chains », *in*: *Data and Applications Security and Privacy XXXIV*, ed. by Anoop Singhal and Jaideep Vaidya, Lecture Notes in Computer Science, Cham: Springer International Publishing, 2020, pp. 24–41, ISBN: 978-3-030-49669-2, DOI: 10.1007/978-3-030-49669-2_2 (cit. on pp. 19, 138).

[10] 5G-PPP, *View on 5G Architecture (Version 2.0)*, 2017 (cit. on pp. 21, 24, 26, 32, 79, 84, 88, 91).

[11] P. Rost et al., « Mobile Network Architecture Evolution toward 5G », *in*: *IEEE Communications Magazine* 54.*5* (May 2016), pp. 84–91, ISSN: 0163-6804, DOI: 10.1109/MCOM.2016.7470940 (cit. on p. 21).

[12] I. Afolabi et al., « Network Slicing and Softwarization: A Survey on Principles, Enabling Technologies & Solutions », *in*: *IEEE Communications Surveys Tutorials* (2018), pp. 1–1, DOI: 10.1109/COMST.2018.2815638 (cit. on p. 22).

[13] X. Foukas et al., « Network Slicing in 5G: Survey and Challenges », *in*: *IEEE Communications Magazine* 55.*5* (May 2017), pp. 94–100, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600951 (cit. on pp. 22, 38, 41).

[14] T. Taleb et al., « On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration », *in*: *IEEE Communications Surveys Tutorials* 19.*3* (2017), pp. 1657–1681, ISSN: 1553-877X, DOI: 10.1109/COMST.2017.2705720 (cit. on p. 22).

[15] Patrick Marsch et al., *5G System Design: Architectural and Functional Considerations and Long Term Research*, 1st Edition, Wiley Publishing, 2018, ISBN: 978-1-119-42512-0 (cit. on pp. 22, 79).

[16] NGMN, *Paper on 5G End-to-End Architecture Framework*, 2017 (cit. on p. 23).

[17] 3GPP, *Specification # 23.501, System Architecture for the 5G System (5GS)*, 2018 (cit. on pp. 23, 24, 26, 30, 53, 59–61, 69, 72, 85).

[18] 3GPP, *Specification # 23.799, Study on Architecture for Next Generation System*, 2016 (cit. on pp. 24, 29, 30).

[19] 3GPP, *Specification # 28.801, Study on Management and Orchestration of Network Slicing for next Generation Network*, 2018 (cit. on pp. 24, 26, 30, 31, 89, 140).

[20] ONF, *TR-526 Applying SDN Architecture to 5G Slicing* (cit. on pp. 24–26, 34).

[21] ETSI, *GR NFV-EVE 012 V3.1.1. Network Functions Virtualisation (NFV) Release 3; Evolution and Ecosystem; Report on Network Slicing Support with ETSI NFV Architecture Framework.* 2017 (cit. on pp. 25, 27, 34, 88).

[22] P. Rost et al., « Network Slicing to Enable Scalability and Flexibility in 5G Mobile Networks », *in*: *IEEE Communications Magazine* 55.*5* (May 2017), pp. 72–79, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600920 (cit. on pp. 25, 39).

[23] 5GAmericas, *Network Slicing for 5G Networks and Services*, 2016 (cit. on pp. 25, 26).

[24] 3GPP, *Specification # 22.891, Study on New Services and Markets Technology Enablers*, 2016 (cit. on p. 25).

[25] Cloudify, *Open NFV Roadmap: ONAP, ETSI, and TOSCA Deconstructed*, 2017 (cit. on p. 27).

[26] Marc Cohn, *New Linux Foundation White Paper: Harmonizing Open Source and Standards in SDN*, 2017 (cit. on p. 28).

[27] NGMN, *Description of Network Slicing Concept*, 2016 (cit. on pp. 28, 29).

[28] 3GPP, *Specification # 28.530, Management and Orchestration; Concepts, Use Cases and Requirements.* 2018 (cit. on p. 29).

[29] 3GPP, *Specification # 22.261, Service Requirements for the 5G System*, 2017 (cit. on p. 34).

[30] K. Katsalis et al., « Network Slices toward 5G Communications: Slicing the LTE Network », *in*: *IEEE Communications Magazine* 55.*8* (2017), pp. 146–154, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600936 (cit. on pp. 37, 38).

[31] J. Ordonez-Lucena et al., « Network Slicing for 5G with SDN/NFV: Concepts, Architectures, and Challenges », *in*: *IEEE Communications Magazine* 55.*5* (May 2017), pp. 80–87, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600935 (cit. on pp. 37, 38, 40, 41).

[32] Misbah Liaqat et al., « Federated Cloud Resource Management: Review and Discussion », *in*: *Journal of Network and Computer Applications* 77 (Jan. 2017), pp. 87–105, ISSN: 1084-8045, DOI: 10.1016/j.jnca.2016.10.008 (cit. on p. 38).

[33] Z. Kotulski et al., « On End-to-End Approach for Slice Isolation in 5G Networks. Fundamental Challenges », *in*: *2017 Federated Conference on Computer Science and Information Systems (FedCSIS)*, Sept. 2017, pp. 783–792, DOI: 10.15439/2017F228 (cit. on pp. 38–40, 110).

[34] X. Li et al., « Network Slicing for 5G: Challenges and Opportunities », *in*: *IEEE Internet Computing* 21.*5* (2017), pp. 20–27, ISSN: 1089-7801, DOI: 10.1109/MIC.2017.3481355 (cit. on pp. 38–40).

[35] S. Vassilaras et al., « The Algorithmic Aspects of Network Slicing », *in*: *IEEE Communications Magazine* 55.*8* (2017), pp. 112–119, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600939 (cit. on pp. 38, 40).

[36] Artur Hecker, *From Slicing to Dynamic Resource Control*, 2017 (cit. on p. 38).

[37] G. Arfaoui, J. M. S. Vilchez, and J. P. Wary, « Security and Resilience in 5G: Current Challenges and Future Directions », *in*: *2017 IEEE Trustcom/BigDataSE/ICESS*, Aug. 2017, pp. 1010–1015, DOI: 10.1109/Trustcom/BigDataSE/ICESS.2017.345 (cit. on pp. 38, 39).

[38] Ashutosh Dutta, *Security Challenges and Opportunities in SDN/NFV and 5G Networks*, 2017 (cit. on p. 39).

[39] Zbigniew Kotulski et al., « Towards Constructive Approach to End-to-End Slice Isolation in 5G Networks », *in*: *EURASIP J. on Info. Security* 2018.*1* (Dec. 2018), p. 2, ISSN: 2510-523X, DOI: 10.1186/s13635-018-0072-0 (cit. on pp. 39, 111).

[40] Danish Sattar and Ashraf Matrawy, « Optimal Slice Allocation in 5G Core Networks », *in*: *arXiv:1802.04655 [cs]* (2018), arXiv: 1802.04655 [cs] (cit. on pp. 39, 111).

[41] Rabia Khan et al., « A Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements, and Future Directions », *in*: *IEEE Communications Surveys Tutorials* 22.*1* (2020), pp. 196–248, ISSN: 1553-877X, DOI: 10.1109/COMST.2019.2933899 (cit. on p. 39).

[42] 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)*, Sept. 2017, pp. 36–41, DOI: 10.1109/CSCN.2017.8088595 (cit. on p. 40).

[43] H. Zhang et al., « Network Slicing Based 5G and Future Mobile Networks: Mobility, Resource Management, and Challenges », *in*: *IEEE Communications Magazine* 55.*8* (2017), pp. 138–145, ISSN: 0163-6804, DOI: 10.1109/MCOM.2017.1600940 (cit. on p. 40).

[44] C. Mannweiler et al., « 5G NORMA: System Architecture for Programmable Multi-Tenant 5G Mobile Networks », *in*: *2017 European Conference on Networks and Communications (EuCNC)*, June 2017, pp. 1–6, DOI: 10.1109/EuCNC.2017.7980662 (cit. on p. 41).

[45] ETSI, *GS NFV-MAN 001 V1.1.1 (2014-12), Network Functions Virtualisation (NFV); Management and Orchestration*, 2014 (cit. on pp. 45, 47, 113).

[46] David F. Ferraiolo, Janet A. Cugini, and David R. Kuhn, « Role-Based Access Control (RBAC): Features and Motivations », *in*: *11th Annual Computer Security Applications Conference*, Dec. 1995 (cit. on p. 53).

[47] V. C. Hu et al., « Attribute-Based Access Control », *in*: *Computer* 48.*2* (Feb. 2015), pp. 85–88, ISSN: 0018-9162, DOI: 10.1109/MC.2015.33 (cit. on pp. 53, 55).

[48] L. Badger et al., « Practical Domain and Type Enforcement for UNIX », *in*: *Proceedings 1995 IEEE Symposium on Security and Privacy*, May 1995, pp. 66–77, DOI: 10.1109/SECPRI.1995.398923 (cit. on pp. 53, 55).

[49] D. E. Bell and Leonard J. LaPadula, « Secure Computer Systems: Mathematical Foundations », tech. rep. MTR-2547-VOL-1, Mitre Corp Bedford MA, Nov. 1973 (cit. on p. 53).

[50] Dorothy E. Denning, « A Lattice Model of Secure Information Flow », *in*: *Commun. ACM* 19.*5* (May 1976), pp. 236–243, ISSN: 0001-0782, DOI: 10.1145/360051.360056 (cit. on pp. 53, 56).

[51] R. S. Sandhu et al., « Role-Based Access Control Models », *in*: *Computer* 29.*2* (Feb. 1996), pp. 38–47, ISSN: 0018-9162, DOI: 10.1109/2.485845 (cit. on pp. 54, 67, 68).

[52] R. S. Sandhu and P. Samarati, « Access Control: Principle and Practice », *in*: *IEEE Communications Magazine* 32.*9* (Sept. 1994), pp. 40–48, ISSN: 0163-6804, DOI: 10.1109/35.312842 (cit. on p. 54).

[53] K. A. Oostendorp et al., « Domain and Type Enforcement Firewalls », *in*: *Proceedings DARPA Information Survivability Conference and Exposition. DISCEX'00*, vol. 1, Jan. 2000, 351–361 vol.1, DOI: 10.1109/DISCEX.2000.825039 (cit. on p. 55).

[54] R. S. Sandhu, « Lattice-Based Access Control Models », *in*: *Computer* 26.*11* (Nov. 1993), pp. 9–19, ISSN: 0018-9162, DOI: 10.1109/2.241422 (cit. on p. 55).

[55] Morrie Gasser, *Building a Secure Computer System*, New York, NY, USA: Van Nostrand Reinhold Co., 1988, ISBN: 978-0-442-23022-7 (cit. on p. 55).

[56] November An Electronic et al., *Secure Computer Systems: A Mathematical Model*, 1996 (cit. on p. 56).

[57] O. Salman et al., « Multi-Level Security for the 5G/IoT Ubiquitous Network », *in*: *2017 Second International Conference on Fog and Mobile Edge Computing (FMEC)*, May 2017, pp. 188–193, DOI: 10.1109/FMEC.2017.7946429 (cit. on p. 56).

[58] Haiwei XUE, « A Multilevel Security Model for Private Cloud », *in*: *Chinese Journal of Electronics* (2014) (cit. on p. 56).

[59] P. Watson, « A Multi-Level Security Model for Partitioning Workflows over Federated Clouds », *in*: *2011 IEEE Third International Conference on Cloud Computing Technology and Science*, Nov. 2011, pp. 180–188, DOI: 10.1109/CloudCom.2011.33 (cit. on p. 57).

[60] Luigi Logrippo and Abdelouadoud Stambouli, « Configuring Data Flows in the Internet of Things for Security and Privacy Requirements », *in*: *11th International Symposium on Foundations and Practice of Security (FPS 2018)* (2018), p. 16, DOI: 10.1007/978-3-030-18419-3_8 (cit. on p. 57).

[61] Uday Tupakula, Vijay Varadharajan, and Kallol Karmakar, « Access Control Based Dynamic Path Establishment for Securing Flows from the User Devices with Different Security Clearance », *in*: *Advanced Information Networking and Applications*, ed. by Leonard Barolli et al., Advances in Intelligent Systems and Computing, Springer International Publishing, 2020, pp. 1303–1315, ISBN: 978-3-030-15032-7, DOI: 10.1007/978-3-030-15032-7_109 (cit. on p. 57).

[62] A. Alshehri and R. Sandhu, « Access Control Models for Virtual Object Communication in Cloud-Enabled IoT », *in*: *2017 IEEE International Conference on Information Reuse and Integration (IRI)*, Aug. 2017, pp. 16–25, DOI: 10.1109/IRI.2017.60 (cit. on p. 57).

[63] Shanay Behrad et al., « A New Scalable Authentication and Access Control Mechanism for 5G-Based IoT », *in*: *Future Generation Computer Systems* 108 (July 2020), pp. 46–61, ISSN: 0167-739X, DOI: 10.1016/j.future.2020.02.014 (cit. on p. 57).

[64] R. Gopi and A. Rajesh, « Securing Video Cloud Storage by ERBAC Mechanisms in 5g Enabled Vehicular Networks », *in*: *Cluster Comput* 20.4 (Dec. 2017), pp. 3489–3497, ISSN: 1573-7543, DOI: 10.1007/s10586-017-0987-0 (cit. on p. 58).

[65] Vladimir Oleshchuk and Rune Fensli, « Remote Patient Monitoring Within a Future 5G Infrastructure », *in*: *Wireless Pers Commun* 57.*3* (Apr. 2011), pp. 431–439, ISSN: 1572-834X, DOI: 10.1007/s11277-010-0078-5 (cit. on p. 58).

[66] M. Pattaranantakul et al., « Leveraging Network Functions Virtualization Orchestrators to Achieve Software-Defined Access Control in the Clouds », *in*: *IEEE Transactions on Dependable and Secure Computing* (2018), pp. 1–1, ISSN: 1545-5971, DOI: 10.1109/TDSC.2018.2889709 (cit. on pp. 58, 82).

[67] 3GPP, *Specification # 28.533, Management and Orchestration; Architecture Framework*, 2018 (cit. on p. 59).

[68] 5G-ENSURE, *Deliverable D2.7 - Security Architecture*, 2016 (cit. on pp. 63, 81).

[69] Ron Ross, Michael McEvilley, and Janet Oren, « Systems Security Engineering: Considerations for a Multidisciplinary Approach in the Engineering of Trustworthy Secure Systems », tech. rep. NIST Special Publication (SP) 800-160 Vol. 1, National Institute of Standards and Technology, Mar. 2018, DOI: 10.6028/NIST.SP.800-160v1 (cit. on p. 67).

[70] NGMN, *Service-Based Architecture in 5G*, 2018 (cit. on p. 69).

[71] 3GPP, *Specification # 23.502, Procedures for the 5G System (5GS)*, 2020 (cit. on p. 72).

[72] A. La Marra et al., « Enhancing Security in ETSI Open Source MANO with Usage Control Capability », *in*: *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*, Apr. 2019, pp. 25–29 (cit. on p. 79).

[73] J. Strassner, *Policy-Based Network Management*, Elsevier, 2004, ISBN: 978-1-55860-859-7, DOI: 10.1016/B978-1-55860-859-7.X5033-6 (cit. on p. 79).

[74] 5G Americas, *The Evolution of Security in 5G*, 2019 (cit. on p. 81).

[75] 5G-PPP, *D2.3, 5G Mobile Network Architecture, Final Overall Architecture*, 2019 (cit. on p. 81).

[76] Borja Bordel et al., « An Inter-Slice Management Solution for Future Virtualization-Based 5G Systems », *in*: *Advanced Information Networking and Applications*, ed. by Leonard Barolli et al., Advances in Intelligent Systems and Computing, Springer International Publishing, 2020, pp. 1059–1070, ISBN: 978-3-030-15032-7, DOI: 10.1007/978-3-030-15032-7_89 (cit. on p. 81).

[77] Daniel Guija and Muhammad Shuaib Siddiqui, « Identity and Access Control for Micro-Services Based 5G NFV Platforms », *in*: *Proceedings of the 13th International Conference on Availability, Reliability and Security*, ARES 2018, New York, NY, USA: ACM, 2018, 46:1–46:10, ISBN: 978-1-4503-6448-5, DOI: 10.1145/3230833.3233255 (cit. on p. 82).

[78] Maxime Compastié et al., « A TOSCA-Oriented Software-Defined Security Approach for Unikernel-Based Protected Clouds », *in*: *2019 IEEE Conference on Network Softwarization (NetSoft)*, June 2019, pp. 151–159, DOI: 10.1109/NETSOFT.2019.8806623 (cit. on p. 82).

[79] T. P. d Souza et al., « SONA: Software Defined Optical Networks Slicing Architecture », *in*: *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, Mar. 2017, pp. 654–661, DOI: 10.1109/AINA.2017.58 (cit. on p. 82).

[80] Estefania Coronado and Roberto Riggio, « Flow-Based Network Slicing: Mapping the Future Mobile Radio Access Networks », *in*: *2019 28th International Conference on Computer Communication and Networks (ICCCN)*, July 2019, pp. 1–9, DOI: 10.1109/ICCCN.2019.8847068 (cit. on p. 82).

[81] Xin Li et al., « Efficient and Secure 5G Core Network Slice Provisioning Based on VIKOR Approach », *in*: *IEEE Access* 7 (2019), pp. 150517–150529, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2019.2947454 (cit. on p. 83).

[82] Eftychia Datsika et al., « Software Defined Network Service Chaining for OTT Service Providers in 5G Networks », *in*: *IEEE Communications Magazine* 55.*11* (Nov. 2017), pp. 124–131, ISSN: 1558-1896, DOI: 10.1109/MCOM.2017.1700108 (cit. on p. 83).

[83] Vijay Varadharajan et al., « A Policy-Based Security Architecture for Software-Defined Networks », *in*: *IEEE Transactions on Information Forensics and Security* 14.*4* (Apr. 2019), pp. 897–912, ISSN: 1556-6021, DOI: 10.1109/TIFS.2018.2868220 (cit. on p. 83).

[84] John Anderson, Matt Piazza, and Aspen Olmsted, « Decentralised, Dynamic Network Path Selection in High Performance Computing », *in*: *2016 International Conference on Information Society (i-Society)*, Oct. 2016, pp. 88–90, DOI: 10.1109/i-Society.2016.7854183 (cit. on p. 83).

[85] Risald, Antonio E. Mirino, and Suyoto, « Best Routes Selection Using Dijkstra and Floyd-Warshall Algorithm », *in*: *2017 11th International Conference on Information Communication Technology and System (ICTS)*, Oct. 2017, pp. 155–158, DOI: 10.1109/ICTS.2017.8265662 (cit. on p. 83).

[86] ETSI, *GS NFV-IFA 014 V2.3.1, Network Functions Virtualisation (NFV) Release 2; Management and Orchestration; Network Service Templates Specification*, 2017 (cit. on p. 88).

[87] 3GPP, *Specification # 28.531, Management and Orchestration; Provisioning.* 2018 (cit. on p. 88).

[88] NIST, *Security and Privacy Controls for Federal Information Systems and Organizations*, 2013 (cit. on p. 90).

[89] Frederic J. de Vaulx, Eric D. Simmon, and Robert B. Bohn, « Cloud Computing Service Metrics Description », *in*: *Special Publication (NIST SP) - 500-307* (2018), DOI: 10.6028/NIST.SP.500-307 (cit. on pp. 93, 118).

[90] Ashish Singh and Kakali Chatterjee, « Cloud Security Issues and Challenges: A Survey », *in*: *Journal of Network and Computer Applications* 79 (Feb. 2017), pp. 88–115, ISSN: 1084-8045, DOI: 10.1016/j.jnca.2016.11.027 (cit. on p. 109).

[91] E Rosen, *RFC 3031, Multiprotocol Label Switching Architecture*, 2001 (cit. on p. 110).

[92] Rui Shu et al., « A Study of Security Isolation Techniques », *in*: *ACM Comput. Surv.* 49.*3* (Oct. 2016), 50:1–50:37, ISSN: 0360-0300, DOI: 10.1145/2988545 (cit. on p. 110).

[93] V. Del Piccolo et al., « A Survey of Network Isolation Solutions for Multi-Tenant Data Centers », *in*: *IEEE Communications Surveys Tutorials* 18.*4* (2016), pp. 2787–2821, ISSN: 1553-877X, DOI: 10.1109/COMST.2016.2556979 (cit. on p. 110).

[94] Rouven Krebs, Christof Momm, and Samuel Kounev, « Metrics and Techniques for Quantifying Performance Isolation in Cloud Environments », *in*: *Science of Computer Programming*, Special Issue on Component-Based Software Engineering and Software Architecture 90 (Sept. 2014), pp. 116–134, ISSN: 0167-6423, DOI: 10.1016/j.scico.2013.08.003 (cit. on p. 111).

[95] J. Gang and V. Friderikos, « Optimal Resource Sharing in Multi-Tenant 5G Networks », *in*: *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2018, pp. 1–6, DOI: 10.1109/WCNC.2018.8377326 (cit. on p. 111).

[96] ETSI, *GS NFV-IFA 026 V3.2.1, Network Functions Virtualisation (NFV) Release 3; Management and Orchestration; Architecture Enhancement for Security Management Specification.* 2019 (cit. on p. 111).

[97] Haipeng Fang, « Method and Apparatus for Managing Network Slices », [Patent] WO201900284, 2019 (cit. on p. 112).

[98] Ruiyue XU, « Method and Device for Deploying Network Slice », [Patent] WO201929268, 2019 (cit. on p. 112).

[99] Xiaoqian Jia, « Service Management Method and Device and Storage Medium », [Patent] WO201929256, 2019 (cit. on p. 112).

[100] Scott Rose et al., « Zero Trust Architecture (2nd Draft) », tech. rep. NIST Special Publication (SP) 800-207 (Draft), National Institute of Standards and Technology, Feb. 2020, DOI: 10.6028/NIST.SP.800-207-draft2 (cit. on p. 141).

[101] Paola Cappanera, Federica Paganelli, and Francesca Paradiso, « VNF Placement for Service Chaining in a Distributed Cloud Environment with Multiple Stakeholders », *in*: *Computer Communications* 133 (Jan. 2019), pp. 24–40, ISSN: 0140-3664, DOI: 10.1016/j.comcom.2018.10.008 (cit. on p. 141).

Acronyms

3GPP 3rd Generation Partnership Project.

5GC 5G Core.

5G-EIR 5G-Equipment Identity Register.

5G-PPP 5G Infrastructure Public Private Partnership.

5GS 5G System.

AAC Authentication and Access Control.

ABAC Attribute-Based Access Control.

ACL Access Control List.

Af Affinity.

AF Application Function.

AMF Access and Mobility Management Function.

AN Access Network.

API Application Programming Interfaces.

AUSF AUthentication Server Function.

BAN Body Area Networks.

BLP Bell-LaPadula.

BSS Business Support Systems.

CapEx Capital Expenditure.

CISI Container Infrastructure Service Instance.

CN Core Network.

CS Communication Service.

CSC Communication Service Customer.

CSG Communication Service Graph.

CSMF Communication Service Management Function.

CSP Communication Service Provider.

DN Data Network.

DTE Domain and Type Enforcement.

eMBB enhanced Mobile Broadband.

EPS Evolved Packet System.

ETSI European Telecommunications Standards Institute.

FCAPS Fault, Configuration, Accounting, Performance and Security.

IaaS Infrastructure as a Service.

IDS Intrusion Detection System.

IoT Internet of Things.

IPS Intrusion Prevention System.

KPI Key Performance Indicator.

LCM Life Cycle Management.

MAC Mandatory Access Control.

MANO Management And Network Orchestration.

MLS Multi Layer Security.

MME Mobility and Management Entity.

mMTC massive Machine Type Communications.

MVNO Mobile Virtual Network Operator.

NAS Non-Access Stratum.

NEF Network Exposure Function.

NF Network Functions.

NFV Network Functions Virtualisation.

NFVO NFV Orchestrator.

NG-RAN Next Generation RAN.

NGMN Next Generation Mobile Networks.

NRF Network Repository Function.

NS Network Services.

NSL Network Slice Links.

NSlice Network Slice.

NSlice Network Slices.

NSliceCh Network Slice Chain.

NSMF Network Slice Management Function.

NSSF Network Slice Selection Function.

NSSMF Network Slice Subnet Management Function.

NWDA NetWork Data Analytics Function.

OAM Operations, Administration and Management.

ONF Open Networking Foundation.

OpEx Operational Expenditure.

OSS Operation Support Systems.

PCF Policy Control Function.

PDU Protocol Data Unit.

PNF Physical Network Functions.

QoS Quality of Service.

RAN Radio Access Network.

RBAC Role Based Access Control.

RDAC Role and Domain Access Control.

RMSE Root Mean Squared Error.

S/PGW Serving GateWay/PDN GateWay User Plane functions.

SaaS Software as a Service.

SBA Service Based Architecture.

SBI Service Based Interface.

SC Security Class.

SDN Software Defined Networking.

SDO Standards Developing Organizations.

SeCOP Security Constraint and Optimization Problem.

SEPP SEcurity Protection Proxy.

SL Security Level.

SLA Service Level Agreement.

SMF Session Management Function.

T Trust.

TCB Trusted Computing Base.

TOSCA Topology and Orchestration Specification for Cloud Applications.

TPM Trusted Platform Module.

UDM Unified Data Management.

UDR Unified Data Repository.

UDSF Unstructured Data Storage Function.

UE User Equipment.

UPF User Plane Function.

URLLC Ultra-Reliable and Low Latency Communications.

V2X Vehicle to Everything.

VANET Vehicular Ad-Hoc Network.

VIM Virtualised Infrastructure Manager.

VL Virtual Link.

VM Virtual Machines.

VNF Virtual Network Functions.

VNFFG VNF Forwarding Graph.

VNFM VNF Manager.

VSF Virtual Security Function.

DOCTORAT BRETAGNE LOIRE MATHSTIC

Université de Bretagne Occidentale

Titre : Sécurisation du slicing dans les réseaux mobiles de 5ème génération

Mot clés : Network Slicing, sécurité, 5G, isolation

Résumé : Le « network slicing » est la pierre angulaire pour la conception et le déploiement de services de communication à forte valeur ajoutée qui seront supportés par les nouveaux cas d'usage introduits par la nouvelle architecture 5G. Ce document souligne le défi que représente l'isolation des « network slices », et la gestion de sa sécurité en fonction des politiques retenues.

Tout d'abord, un nouveau modèle de contrôle d'accès a été créé. Il permet de sécuriser les interactions entre les fonctions réseaux supportées par les systèmes 5G. Ensuite, la gestion des interactions entre les « network slices » a été abordée. On utilise le concept de chaînes de « network slices », qui seront mises en oeuvre après validation des contraintes de sécurité selon la politique choisie. Enfin, une méthode de quantification de l'isolation a été mise au point, permettant de connaître le degré d'isolation d'un service de communication offert via des « network slices ». Cela permet aux opérateurs de réseau et aux clients de mesurer le degré d'isolation, puis d'améliorer la configuration des « network slices » afin de le renforcer.

Ces éléments établissent un cadre solide contribuant à sécuriser, verticalement, les services de communication d'un réseau 5G et à évaluer le degré de sécurité en ce qui concerne leurs interactions et leur isolation.

Title: Securing network slices in 5th generation mobile networks

Keywords: Network Slicing, security, 5G, isolation

Abstract: Network slicing is a cornerstone in the conception and deployment of enriched communication services for the new use cases envisioned and supported by the new 5G architecture. This document makes emphasis on the challenge of the network slicing isolation and security management according to policy.

First, a novel access control model was created, that secures the interactions between network functions that reside inside the 5G system. Then, the management of the interactions between network slices was addressed. We coin the concept of network slice chains, which are conceived after security constraint validation according to policy. Lastly, a method to quantify isolation was developed, permitting to find out how well isolated a communication service is, which is offered via network slices. This enables network operators and customers to measure the isolation level and improve the configuration of the network slices so the isolation level can be enhanced.

These components establish a solid framework that contributes to secure, vertically, the communication services of a 5G network and assess how secure they are with respect to their interactions and isolation.