Kai Lampi

5G CORE NETWORK SLICING

MEC and NEF security concerns while exposing the 5G core for 3rd parties

Master's thesis

Cybersecurity

2021

South-Eastern Finland
University of Applied Sciences

Author (authors)	Degree title	Time
Kai Lampi	Master of Engineering, Cybersecurity	September 2021

Thesis title

5G core network slicing
MEC and NEF security concerns while exposing the 5G core for 3rd parties

86 pages
4 pages of appendices

Commissioned by

Elisa Oyj, Eetu Prieur

Supervisor

Vesa Kankare

Abstract

The objective of the thesis was to discover 5G core network architecture readiness to provide network slicing service. The 5G core slicing is the base service to support mobile edge computing (MEC) in the 5G network. In order slicing to be dynamically provisioned, network expose function (NEF) is needed. These all are new services and their influence over the 5G core network security is not well known. Information available was theoretical, but not practical.

The theory part relies on a systematic literature review. A qualitative method was chosen because there was nothing to measure. Interviews were carried out with case study methods and a pre-defined theme interview structure. Finally, the interview answers were compared to the literature review and iteration made to complement the literature review part.

The 5G core, called as service-based architecture (SBA) is needed for full end-to-end slicing. Mobile edge computing MEC is needed to provide minimal delays between the mobile device and service, for example for autonomous driving. MEC has several different deployment models, and they provide different balance between delays and security. Network exposure function (NEF) provides the interface to allow 3rd parties to create slices in the mobile network and carry out future services that end users can purchase from other service provider than the mobile operator.

The study showed that SBA core is ready to be implemented. Encryption is available by default and that makes eavesdropping and other traditional hacking methods hard to accomplish. MEC protection requires mechanisms to control what is allowed to run on it and several firewalls create security zones. NEF is placed in the edge of the operator's network and exposed to public network. It requires access control to limit unwanted authentication requests. The overall security control presumes situational awareness system and AI- and ML-based security solutions to adapt quickly changing traffic patterns.

Keywords

5G, core network, slicing, cybersecurity, service based architecture

CONTENTS

1 INTRODUCTION

5G is the latest generation of the land mobile telecommunication, defined by the ITU's 3GPP (3rd Generation Partnership Project) standard continuum. The first commercial 5G network was launched in Finland June 2018 (Lehto 2018). 5G is said to change everything. It provides better connectivity everywhere and enables services we cannot even image yet. 5G brings faster transfer speeds, better coverage and will allow doing everything remotely as well as nowadays locally. Remote surgery, remote manufacturing, remote driving, anything that can be done remotely will be possible remotely. The world will be virtualized, and it can be experienced over virtual reality glasses and with augmented reality. That all should save time and environments. (Traficom 2021).

Consumers are especially interested in 5G's increased data transfer, but also lower latencies. A higher data rate makes almost any service to work smoother than at a lower rate. For example, upcoming 4K and 8K videos at a high frame rate in cloud gaming are dependent on both high data rate and low latencies. Fixed wireless access with reasonable and reliable rates that can replace old copper-wire access is another example of new 5G services requiring high transfer rates.

Lower latency improves mobile gaming experiences, but it also brings totally new services available. The business sector is looking for a reliable mobile network to use it for autonomous driving or power grid control type of applications. In that case low latency and ultra-reliability need to be combined.

In the early phase, 5G technology was concentrated to provide faster data rates. Faster speed has also been the strong 5G story. However, 5G should be comprehensive network for everything. Imagine if the mobile network is full of consumer customers' usage and it should also serve industrial robots' control systems or thousands of Internet of things (IoT) sensors. How to guarantee that each of them can get the service they need? Luckily, we have seen only the first of three main traffic categories 5G is going to offer and these are just basic beginning categories. These categories are listed below.

1. Enhanced Mobile Broadband (eMBB), targeted for high data rate and high data traffic.
2. Ultra-reliable and Low Latency Communications (URLLC), targeted for low latency, low error rate and ultra-reliability.
3. Massive Machine Type Communications (mMTC), targeted for large numbers of connected devices and energy saving.

(Ta-Hao et.al. 2019)

Enhanced mobile broadband aims to provide almost fiber speeds over the air and its benefits are easiest to understand for everyone. It improves especially radio path with new base stations and end devices (phones, mobile routers). Ultra-reliable and Low Latency Communications (URLLC) and Massive Machine Type Communications (mMTC) are both for machine-to-machine communications and their benefits are more difficult to understand for large audience. The mMTC is to provide connections for large amount of different kind of IoT-sensors, like temperature, humidity, pressure, movement, vibration, magnetic and location. All of that work with minimal power consumption and maximize battery life. At the mobile network side mMTC means that up to 1 million sensors can be located at one square kilometer area. The URLLC is needed especially for autonomous driving and other solutions that need reliable and low latency communication, like power grid control. (ENISA 2019.)

To enable 5G to efficiently serve all these different kinds of requirements, slicing was developed. It means that 5G network capacity can be partitioned and not everything is served in the best effort style. Some traffic has higher transfer speed than another, and some traffic has lower delays than another. Network slicing is a key to provide the above-mentioned services in a resource-efficient way in the mobile network. Both radio path and core network can be divided into multiple slices. Each slice can have its own characteristics and resource allocation. Slicing can provide and guarantee certain capacity, which leads to real quality of service (QoS) for certain customer or traffic type. Efficient QoS is something that has not been available before 5G.

Slicing and low delays are logical to be bundled together, as many services requiring low delays also require guaranteed capacity, ie. QoS. However, all three main categories, eMBB, mMTC and URLLC, are designed to be sliceable features, as they require different delays, power consumption and throughput. Figure 1 presents simple overview how sliced services are set on top of the physical network, and how the physical layer is divided into three parts: access, transport and core network. In this figure, the core network part is what this master's thesis concentrates on. Even though radio slicing is the most apparent part of the 5G slicing, most important functionalities from the cybersecurity point of view are happening on the core slicing side. Radio security does not change if the radio access is sliced or not, because there are still same authentication mechanisms and encryptions. When slicing is deployed in the network, numerous things change in the core side. All slicing definitions are configured into the core side network functions, and everything is controlled by the core systems. Especially core network isolation changes, when a gate for third parties is added to provision slices. That is what core slicing is about: to provide new kind of services in the mobile core network. This demands cybersecurity considerations.

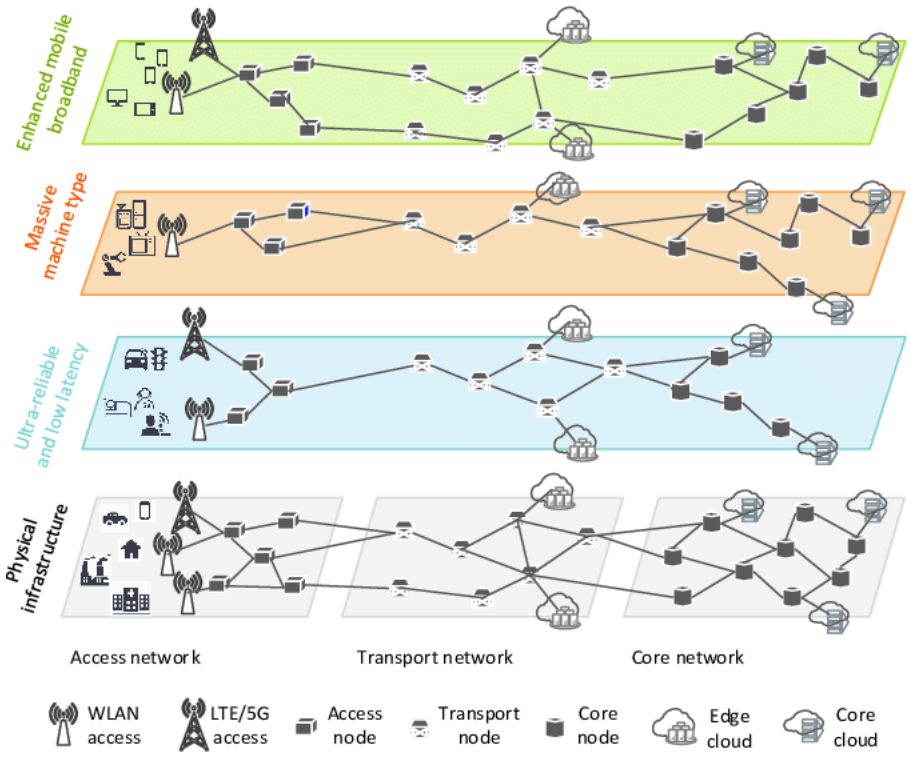


Figure 1. 5G Slicing overview. (Guan et.al. 2018).

1.1 5G releases

5G is not just one single standard, but it consists of many different releases. Each release defines a new set of features and 5G evolves through releases. The release development begins with proposed features, that may or may not be accepted as study items. This phase is called as package approval. Accepted study items are developed and defined in Technical Specification Groups (TSG). TSGs work has three stages to freeze until the whole release is freeze. The freezing means that all necessary features are specified and vendors can finalize their products according it. If TSGs' work meets delays, they can either postpone the freeze or postpone features to next release. (3GPP FAQs N.d.)

A mentioned above, the technical specification development work within 3GPP is accomplished by Technical Specification Groups (TSGs). There are three TSGs, TSG radio access network (TSG RAN), TSG service and systems aspects (TSG SA) and TSG core network & terminals (TSG CT). Each of them has one to six workgroups (WG) to oversee more specific technology part, like TSG RAN WG1 for physical radio layer, TSG SA WG2 for system architecture and services or TSG CT WG4 for core network protocols. For the topic of this master thesis relevant TSGs are Service and System Aspects (TSG SA) that is responsible for the overall architecture and service capabilities of systems based on 3GPP specifications and Core Network and Terminals (TSG CT) that is responsible for specifying terminal interfaces, terminal capabilities, and the core network part of 3GPP systems. (3GPP Specifications Groups 2021.)

The 5G phase 1 specification was 3GPP release 15 and the current release is 16. The release 17 is expected to be freeze i.e., ready in Q3/2022 (3GPP Release 17 timeline agreed 2020). Rel-15 contained all the major specifications needed for the 1st phase 5G service, including the following:

- NR "New (5G) Radio"
- architecture for Non-Stand Alone (NSA) core. This means that New Radio (NR) is used with the LTE core network (EPC)
- architecture for Stand Alone (SA) core. Means that NR is connected to 5G core network (5GC)

- Service-Based Architecture (SBA) for 5G core
- support for edge computing
- network slicing (logical end-2-end networks)
- policy framework and QoS support
- network capability exposure
- Multi-Operator Core Network (MOCN)

(3GPP Release 15.)

Rel-16 contains many additions to the radio part, and many of them are specifically aimed at different industry verticals. (Putkonen 2019.) In addition, the network slicing function is being further improved, providing better tools to authentication and authorization per network slice. Another improvement is enabler for Network Automation (eNA), which, for example, can provide slice load level information to other 5G core elements (5G Americas 2021, 54). Rel-17 contains mainly enhancements for previous releases features. For example, the subjects of this thesis, network slicing and edge computing, are receiving improvements (3GPP Release 17 2020). This thesis network functions are dealing with implementation ready features, which means that they are according to the Rel-15.

5G Network Slicing, Mobile Edge Computing (MEC) and Network Exposure Functions (NEF) together enable mobile network operators (MNO) to provide certain constant capacity and edge computing to (enterprise) customers to gain remarkably lower delays in mobile communication than in the past. It is also a new situation for MNOs that they can (and to succeed in the business, they must) expose their mobile network for 3rd parties. This makes new kind of services and business models possible, but on the other side, opens new threat landscapes. This master's thesis will be concentrating on 5G network slicing, granting 3rd party access to 5G services, their threat landscapes, and ways to mitigate risks.

2 RESEARCH PROBLEM AND METHODS

This master's thesis is about to study 5G core network slicing, Mobile Edge Computing (MEC), and Network Expose Function (NEF) from the cybersecurity point of view. The topic is relevant and justified, because they are new functions

that did not exist as such in the 4G era and the number of the research studies of the 5G core slicing topic are limited. A majority of the 5G slicing research focuses on 5G radio part slicing.

The idea of this thesis is to study how those functions are defined in the 5G standard at the general level, how a mobile network operator (MNO) would implement them, and what kind of services they can offer to customers. One of the goals is also improve the researcher's own and his organization's awareness of the topic. The researcher will do conclusions on the threats and mitigation methods, but implementation of the methods is not a target.

To get comprehensive view end-to-end 5G network slicing, 5G Radio Access Network (RAN) should also be covered. However, to keep this master's thesis in the required level of details and focused on the cybersecurity, it is mandatory to delimit all RAN parts out. This thesis covers only the 5G core, and RAN parts are covered only if necessary.

2.1 Research problem

Until 5G, a mobile operator's resource allocation has been static and mobile core has been closed system. Only certain interfaces have been open to other operators to allow roaming and billing type of functions, but the threat landscape have been limited. 5G aims to be more flexible, and as mentioned in the introduction, 5G is designed to serve efficiently several different traffic types. To accomplish that, mobile operators are going to implement network slicing and MEC. They are new techniques with a minor practical experience in production networks. They open new revenue models for MNO in two ways. Slicing itself is waited by business customers, they could obtain dedicated capacity. The other way is new business partners, who can develop own mobile services that utilize mobile network operator's (MNO) resources. In the latter case MNO opens its core network to third party totally new way.

It is inevitable that MNO will increase its mobile core network threat landscape when the mobile core interfaces will be opened to 3rd parties. From the security

point of view purely, it is even worse that 3rd parties can dynamically allocate resources, theoretically it opens a new denial-of-service (DoS) attack vector. The research problems are that MNO should utilize new technology to provide new services and grow its business. This also means that new attack vectors are opened, and new kind of threats arisen. The dilemma is how to prevent abuse, for example over allocation, but provide new 5G services in a secure way.

2.2 Research questions

The primary question is:

1. What does 5G network slicing mean for the operator's core network security?

Two secondary questions are:

2. What kind of new threats will arise?
3. What actions should be considered to mitigate the risks?

2.3 Research methods

The major method decision was made between qualitative and quantitative methods. That choice was the qualitative, because

1) qualitative research information is based on evaluation, data sources are textual and observational
2) quantitative research presumes measurable data and there is no data to measure.

(Baškarada 2013).

This master thesis consists of two parts: 1) a literature review and 2) a qualitative case study research. The key factor in the case study research project is the definition of the research problem. Well formulated research questions and study objectives presume that every case study should begin with a comprehensive literature review (Baškarada 2013). That is the reason why this master thesis as well includes literature review part. The case study was chosen, because it is a method for learning about a complex instance, based on a comprehensive understanding of that instance (Baškarada 2013). Others reinforce that:

Robert Yin's definition for the case is:

> a contemporary phenomenon within its real-life context, especially when the boundaries between a phenomenon and context are not clear and the researcher has little control over the phenomenon and context. (Yin 2002, 13.)

Robert Stake's definition is:

> a specific, a complex, functioning thing," more specifically "an integrated system" which "has a boundary and working parts. (Stake 1995, 2.)

Data gathering methods in the qualitative case study are observation, interview and document review (Yazan 2015) i.e., they seem to fit well to 5G core network, because it is a complex set of devices, software and configurations, but there is nothing to measure in this case. The obvious outcome is that this this study should be implemented as a qualitative case study. That was the decision even though a qualitative case study has been said to be soft method, meaning that it is easy and not particularly rigorous. On the other hand, it is also said to be remarkably difficult to execute well in practice (Baškarada 2013).

Expert interviews were essential to bring theoretical, standards and white paper type knowledge closer to the practice. A full-scale operator grade 5G core network would be impossible to set up for studying and therefore this type of research would be only a literature review without interviews. Benefits of the interviews in this master's thesis were that researcher can ask further questions to clarify the answers and phenomenon (Hirsjärvi et.al. 2015, 35). In a case study research, there is no exact number of interviewees to be considered as enough. The number can be even one. (Hirsjärvi et.al. 2015, 58.) In this thesis the number of interviewees was five. They were selected by their expertise background.

There were no exact questions to answer, but an interview theme that worked like a checklist in the interview. Themes are the main categories and questions are to be conducted from them. It is characteristic of a semi-structured interview

that not only the interviewer is conducting and clarifying questions, but the interviewee also is an active party (Hirsjärvi et.at. 2015, 66). The active party means that interviewee can introduce topics and relations that are important to the theme, but which possibly were not asked by the researcher. The interview theme is in Chapter 6.

3 5G CORE NETWORK

The world has evolved from fixed dedicated servers towards Software as a Service (SaaS), Platform as a Service (PaaS), etc. for years. Around 2015 Software Defined Wide Area Network (SD-WAN) and Software Defined Local Area Network (SD-LAN) started to take over place from traditional Multiprotocol Label Switching (MPLS) and Ethernet switch-based networks. The same evolution has been a goal in the mobile networks as well. The 4G LTE core was constructed of fixed-function, hard wired and appliance-based architecture. (Ivezic 2020.) In the 5G, a development goal has been that 5G fulfills all these properties: flexibility, programmability, reliability, resilience, multi-tenancy support, isolation and cost-effective resource consumption. (Sun et.al. 2019). This makes 5G core (5GC) to be called as Service Based Architecture (SBA). It means that 5G core is constructed of separate interconnected Network Functions (NF), which are authorized to access each other's services (Ivezic 2020). 5G network is also designed to expose some of its resources to 3rd parties over the Network Expose Function (NEF) to accelerate creative services based on mobile operator services.

The 5GC consists of two major planes, separated by function type, and carried data type. They are named quite logically. User Plane (UP) carries user data and Control Plane (CP) carries all the control data. When the 5G architecture is presented as functions and separated by planes, it is easy to understand different services and their relations as presented in figures 2 and 3.

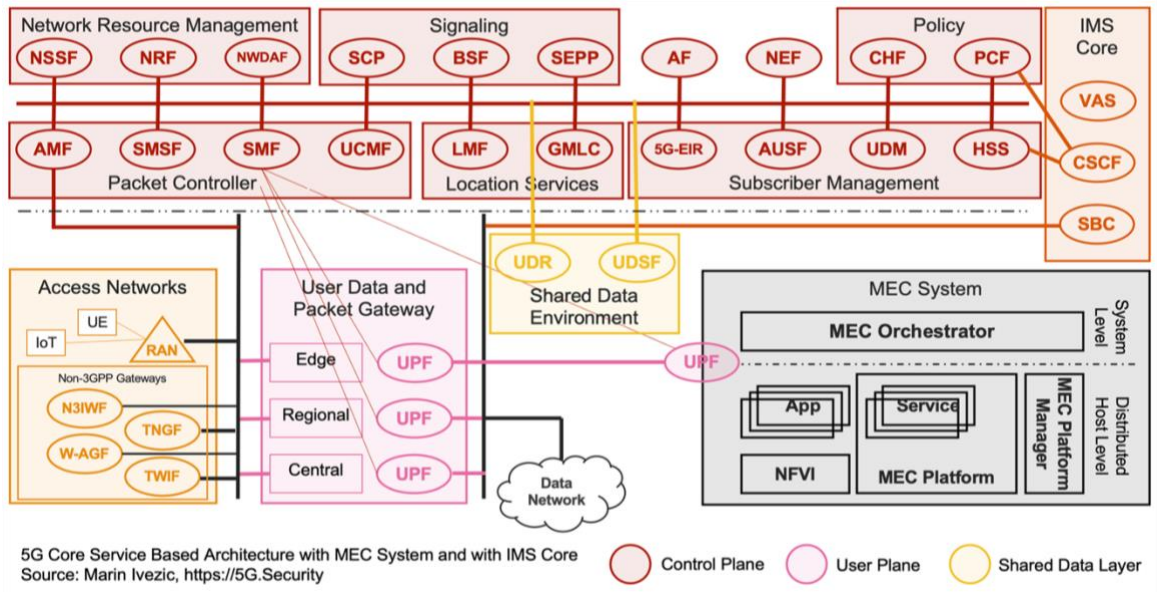


Figure 2. 5G core SBA functions. Access Networks also presented in the figure, even though it is not part of the core, but Radio Access Technology (RAT). (Ivezic 2020).

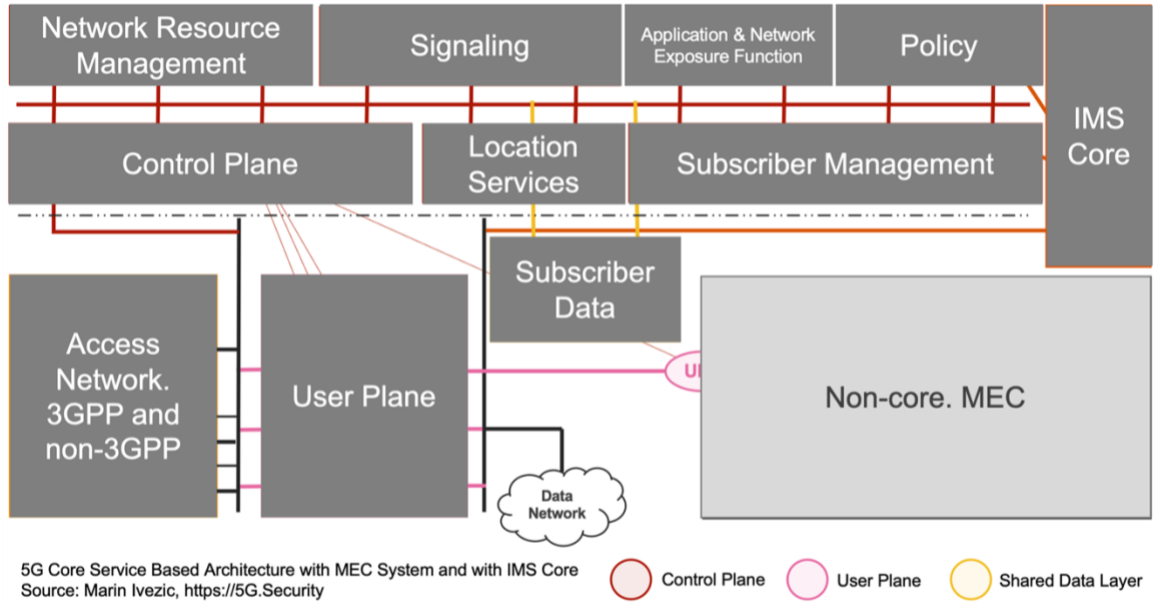


Figure 3. The 5G Core presented as logical blocks. (Ivezic 2020).

3.1 Brief Explanation of 5GC Functions

Network functions enable access and communication between services. Each network function can act as a service producer, but also as a service consumer when it needs a service from another network function. Depending on the function they are Control Plan (CP) or User Plan (UP) functions or some of them can also span the layers. In the SBA network functions can be either physical or virtualized resources. (Golic et.al. 2018.)

Network Resource Management block contains three functions:

1. Network Repository Function (NRF) serves as a repository of the services. It allows every network function to discover the services offered by other network functions. This means also a discovery mechanism that allows 5G elements to discover each other, which enables status updates of the 5G elements.
2. Network Slice Selection Function (NSSF) assigns the Network Slice Instance (NSI) and redirects traffic to a network slice. The selection is based on information provided during user equipment (UE) attach. Finally, a set of Access and Mobility Management Function (AMF) are provided to the UE based on which slices the UE has access to.
3. Network Data Analytics Function (NWDAF) is responsible for providing network analysis information from other network functions.

Signaling block contains three functions:

1. Security Edge Protection Proxy (SEPP) is used in a roaming case. It protects control plane traffic that is exchanged between different 5G operator networks.
2. Service Communication Proxy (SCP) consist of control and user plane. It provides routing control, resiliency, and observability to the core network. To support that, the deployment needs to be alongside of 5G Network Functions (NF), which means SCP to be a decentralized solution.
3. Binding Support Function (BSF) is used for binding an application-function request to a specific Policy Control Function (PCF) instance. It can be compared to LTE network's Policy and Charging Rules Function (PCRF) binding function which was provided by Diameter Routing Agent (DRA) and user for Voice over LTE (VoLTE) or Voice over WiFi (VoWiFi).

Application Function and Network Exposure Function contains two functions:

1. Network Exposure Function (NEF) is used to expose APIs from/to external systems. It is like a gateway to securely expose the services and capabilities of the operator mobile core network functions.
2. Application Function (AF) supports application impact on traffic routing, accesses NEF and interacts with policy framework for policy control.

Policy contains two functions:

1. Policy Control Function (PCF) main purpose is to control the 5G network behavior by supporting a unified policy framework. It is responsible to retrieve subscription information for policy decisions made by the User Data Repository (UDR). Another important role is to support the new 5G QoS policy and charging control functions.
2. Charging Function (CHF) is to allow charging services to be offered to authorized network functions.

Packet Controller contains four functions:

1. Access & Mobility Management Function (AMF) supervises authentication, connection, mobility management between network and device. It receives connection and session related information from the UE.
2. Session Management Function (SMF) covers session management, IP address allocation, and control of policy enforcement.
3. Short Message Service Function (SMSF) supports text message (SMS) transfer over the non-access stratum (NAS).
4. UE radio Capability Management Function (UCMF) is to storage of dictionary entries corresponding to either PLMN-assigned or manufacturer-assigned UE Radio Capability IDs.

Subscriber Management contains four functions:

1. Authentication Server Function (AUSF) is located in the home network and performs authentication functions. It relies on backend service authenticating data and keying materials when 5G Authentication and Key Agreement (5G-AKA) is used. 5G-AKA is one of the techniques available in 5G for mutual authentication between the subscriber and the network. AUSF provides authentication functions of the Home Subscriber Server (HSS). HSS contains user- and subscriber-related information.
2. Unified Data Management (UDM) is a converged repository of subscriber information. It is used to serve several network functions. The 5G UDM can use the User Data Repository (UDR) to store and retrieve subscription data.
3. Equipment Identity Register (5G-EIR) enables authentication of devices in the network. It protects networks and billing against the use of stolen and unauthorized devices.
4. Home Subscriber Server (HSS) is similar to LTE network's HSS. It means that customer profile data and authentication information along with encryption keys are stored in the HSS

5G Location Services contains two functions:

1. Location Management Function (LMF) main task is to determine the location of the end device and provide that information to other NFs. It does not use any satellite system but obtain downlink location measurements or a location estimate from the UE, obtain uplink location measurements and non-UE associated assistance data from the 5G Radio Access Network (RAN).
2. Gateway Mobile Location Center (GMLC) is used to exchange UE's location information also in the roaming case. When the location information is requested by the end device, the GMLC sends location service request to Access & Mobility Management Function (AMF). An emergency call is an example where GMLC is applied. In the emergency call the AMF initiates the location request to the LMF. The LMF processes the location services request, next the LMF returns the result of the location back to the AMF and finally the AMF returns the location service

result to the GMLC, which can then provide the end device location information for the emergency call center agent.

(Ivezic 2020).

3.2 5G core network slicing

The term slicing is used for both the core network and the radio access network. This thesis focuses on the core network slicing only. Network slicing is to divide single physical network into multiple logical virtual networks and can provide service level agreements (SLA) for different needs. For example, services like smart-parking meters value high reliability and security, but are more forgiving with respect to latency, others like driver-less cars may need ultra-low latency (URLLC) and high data speeds. The 3GPP has recognized network slicing to be an essential overall component of the 5G.

A slice in the core network consists of a group of network functions (NFs) that are required for the slice services. Those NFs can be exclusively reserved to that slice only or be shared among multiple slices. The network functions can be either physical or virtualized. One physical node may host several NFs, depending on capacity and security requirements. A shared NF can provide services to several slices. (AdaptiveMobile Security 2021.)

It is practical and cost-efficient that network slicing is virtualized. In that case it is based on network functions virtualization (NFV) and Software Defined Networks (SDN). NFV provides the basis for placing various network functions in different network perimeters and eliminates the need for function or service-specific hardware. SDN complements NFV and they enable the 5G network infrastructure to share the same resources for multiple use cases. (Ahmad et.al. 2019.) Figure 4 illustrates how SDN defines networks at the infrastructure resources layer (the blue layer). NFV assigns functions at the business enablement layer (the green) and on top of them are applications. The slice will be created vertically over them.

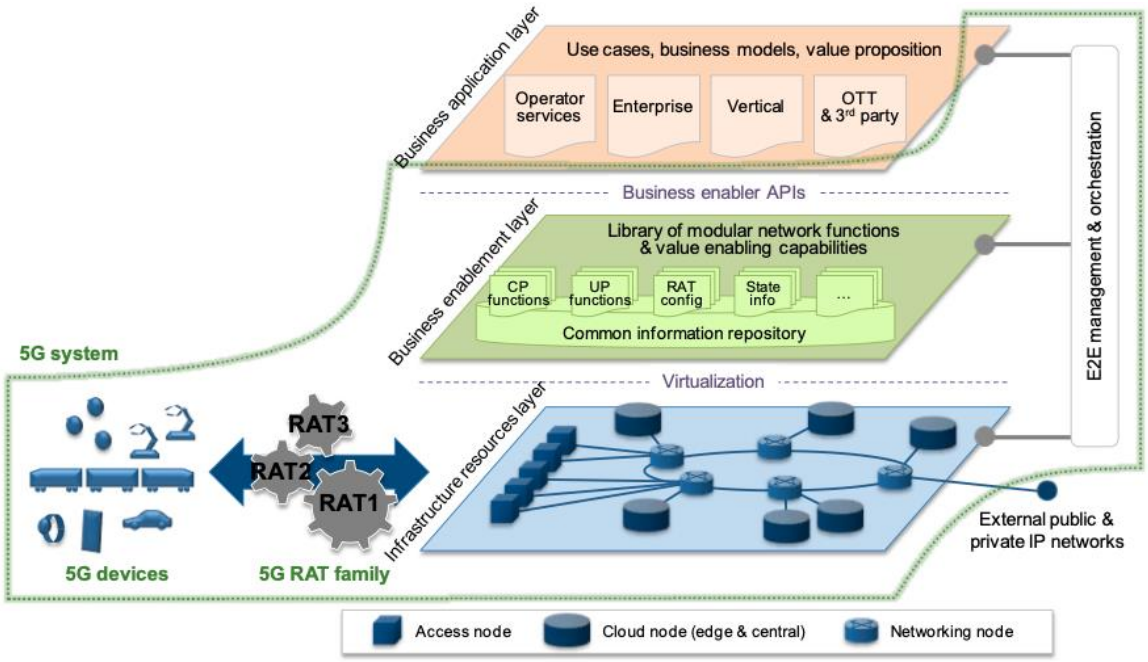


Figure 4. 5G architecture as layer model thinking (NGMN Alliance 2015).

Slices are logical groups of network functions required to fulfill a business requirement. The business requirement then defines the needed capacity and quality features of the slice, meaning transfer rates and delay requirements. The 3rd Generation Partnership Project (3GPP) currently defines in TS 23.501 the following three types of network slice types, based on their quality-of-service features:

1. enhanced Mobile Broadband (eMBB)
2. Ultra-Reliable Low Latency Communications (URLLC)
3. massive Machine Type Communication (mMTC)

Massive Machine Type Communication (mMTC) covers slices that serve a large number of Internet of Things (IoT) devices. It is typical that data amounts are small, some measurement value every now and then, and other times there is no traffic at all. This kind of traffic does not need high bandwidth. It is a radio part topic, but when considered IoT sensors, they are often needed to be placed in weak radio signal locations, like basements. One of the special characteristics of mMTC slice is that it provides great coverage in the cost of transfer rate. Because of low transfer rates and rare communications periods, a huge number of devices can be located at the same area, even 1 million devices per square kilometer.

Enhanced Mobile Broadband (eMBB) slice is aiming to provide as high transfer rate as possible. The radio network part has major role to fulfill this slice type requirements. Still unlaunched (mid-2021) millimeter wave (mmW) frequencies are needed to provide gigabit-class speeds as general availability. Very capable eMMB might also require user plane (UP) data to have local / regional break-out from the transmission network without requiring UP data to travel to the central site. Typically, eMMB is used to serve entertainment use cases, like event streaming.

Ultra-Reliable Low Latency Communications (URLLC) is for mission-critical networks. As its name states, the purpose is to combine two characteristics to provide slice needed for different kind of high accuracy remote control applications. The ultra-reliable part is mainly radio network demand, but low latency puts requirements on the core side as well, especially for the multi-access edge computing (MEC).

It is possible that currently defined slice types may evolve to mixed model. For example, a combination of eMBB's transfer rates and URLLC's low latency might be useful for online gaming, or URLLC might be split into two - one for reliable transfers and one for low latency applications.

3.3 Multi-access Edge Computing (MEC)

Multi-access edge computing (MEC) facilitates 5G's low latency services by bringing computation and storage near to end-user devices. The initial requirement by 3GPP for 5G edge computing was to support low latency together with mission critical and future IoT services. Applications running on edge computing servers offload UE traffic from the core network. The acronym MEC stands for Multi-access Edge Computing defined by ETSI 3GPP. In some context former Mobile Edge Computing can be still seen. (Kekki et.al. 2018.)

Together with 5G network slicing, MEC is a key component of the Massive IoT. A massive number of sensors may produce a huge amount of data that can be processed by MEC near the data originate location and without a need to transfer

data to datacenter first. The network slicing allows offering dedicated network and edge computing resources for service tenants and specifically tailored to their needs. (Kekki et.al. 2018.)

Another basic example of efficient edge computing application is a location application. Traditional way has been to deploy Location Management Function (LMF) in the central core. The user equipment (UE) location has been stored in there and any application requiring that information has retrieved it from the central core. In the 5G SBA architecture, the location service can be deployed in the MEC and located at the edge network. Any application hosted in the MEC can get the UE location information quickly over the short path, as shown in Figure 5. Generally, SBA offers the choice to distribute control plane services (CPS) and user plane services (UPS) in an optimal location to support low latency requirements. (Sun et.al. 2019.)

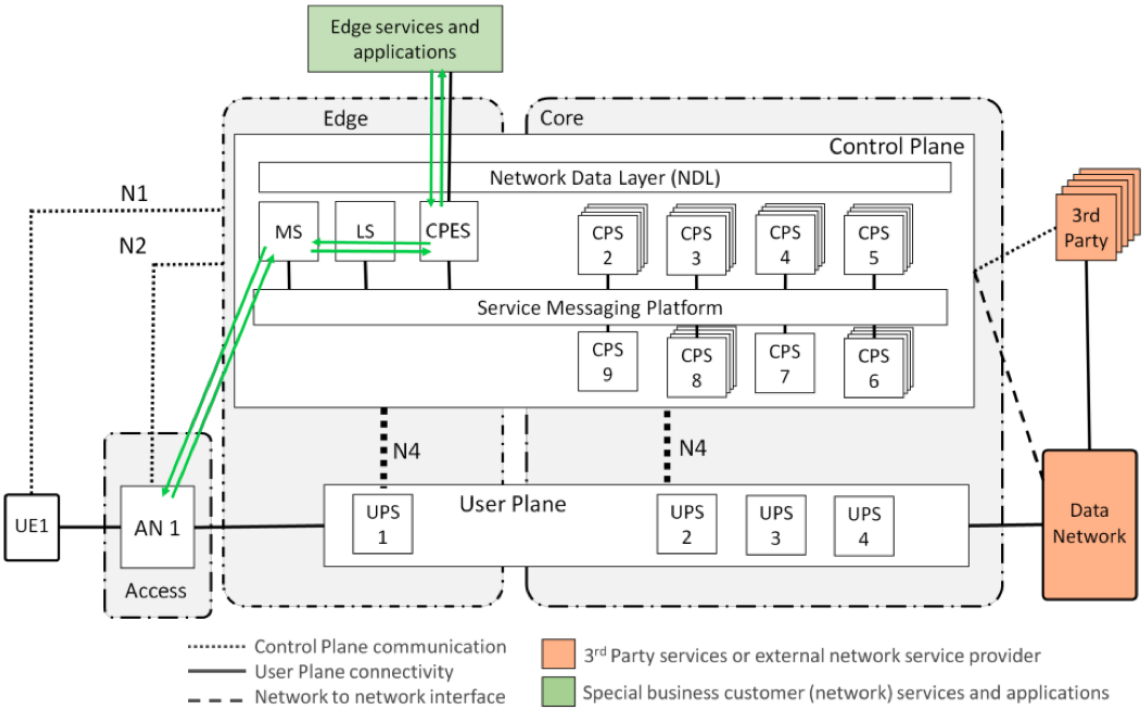


Figure 5. 5G SBA, MEC example. Any application hosted in the edge can get the UE location through the green line (Sun et.al. 2019).

There are multiple options how to physically deploy MEC hosts. The best choice depends on operational, performance or security related requirements. The

following Figure 6 gives an outline of four alternative options for the physical placement of the MEC.

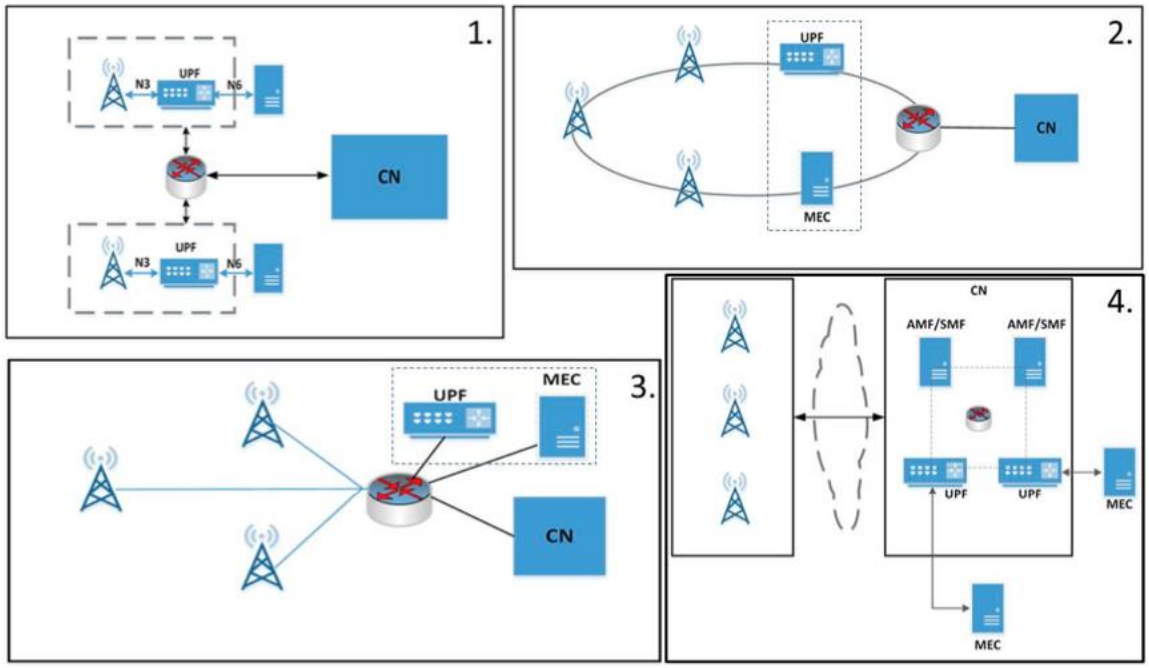


Figure 6. Examples of the physical deployment of MEC (Kekki et.al. 2018).

In Figure 6, the following location alternatives are presented:

1) MEC and the local user plane function (UPF) collocated with the Base Station.
2) MEC collocated with a transmission node, possibly with a local UPF
3) MEC and the local UPF collocated with a network aggregation point
4) MEC collocated with the Core Network functions (i.e. in the same data center)

The first option provides the shortest delays between the UE and MEC. Downsides are that each base station requires a MEC placed next to it to fulfill shortest possible delays. Also, physical security might be weak. The second option would be a feasible compromise with low delays, but one MEC serving several base stations. The third option differs slightly form the second one. Mainly it can support more traffic load because it is not collocated with a transmission node. The fourth deployment model can support heavy loads, but when located at a data center, it cannot provide very low delays.

3.4 Network Exposure Function (NEF)

Before Network Expose Function can be described and understood, the relation between network infrastructure, network service and functions and network management need to be opened. Figure 7 simplifies their relations.

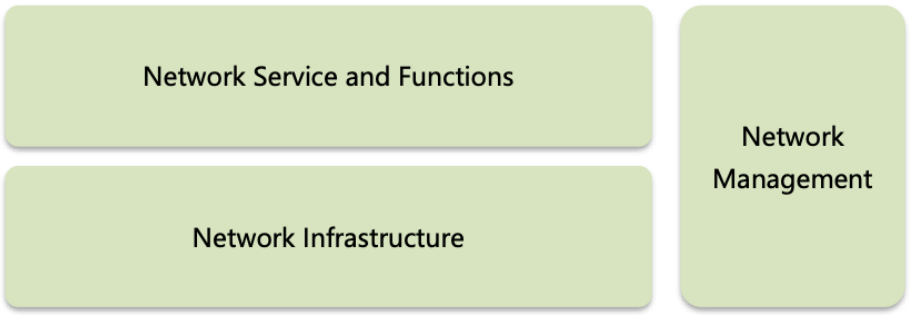


Figure 7. Network capabilities categories (Golic 2018).

Network Service and Functions (NFS) provides access and communications services to users. Functions may include both 5GC control plane and user plane.

Network Infrastructure provides physical or virtual resources for NFS. Physical resources can be servers, access nodes, cloud nodes, networking nodes and associated links (see Figure 2 above for the concept of the node). Virtualized resource can be Virtual Machines (VM), containers, virtualization management software, software platforms, operating system and virtual links. (Golic 2018).

Network Management is used to manage services and functions. Future plans are that network management defines network slices to be used for a given application scenario, chains the relevant modular network functions, assigns the relevant performance configurations, and finally maps all of this onto the infrastructure resources (Golic, 2018). Once there are so many tasks performed by 5GC network management, it is easy to understand why it should be as much automated as possible and why automated network operation has high value in 5G standardization work. In the future network management can be automated and it translates use cases and business models directly into actual network functions and slices.

Exposure of network capabilities means that MNO provides access to its network resources to 3rd party. Resources can be functionalities of the network services and functions, network infrastructure and network management. This allows the 3rd party business partner to provide their own services to their own customers. Network Expose Function (NEF) is the function used to exposure those capabilities. (Golic 2018.)

The granted access level for the 3rd party defines exposure scenarios. Each scenario permits different level access to operator network resources. The 3rd party will have different administrative domains and impact on the network while the capability exposure needs to be maintained in control of MNO all the time. The following three levels of exposure scenarios can be identified. (Golic 2018).

Level 1: Passive exposure – The 3rd party has only passive access to exposed network service and functions. It is not allowed to change, control or manage the exposed network capabilities. The 3rd party can provide input data to the network and obtain the corresponding output data from the network. Also, the 3rd party can request and obtain some data from the network (e.g., network traffic data). Passive access to exposure can be implemented by a NEF interacting with other NFs via APIs. (Golic 2018.) Figure 8 presents that administrative domains are separate in the passive exposure case.

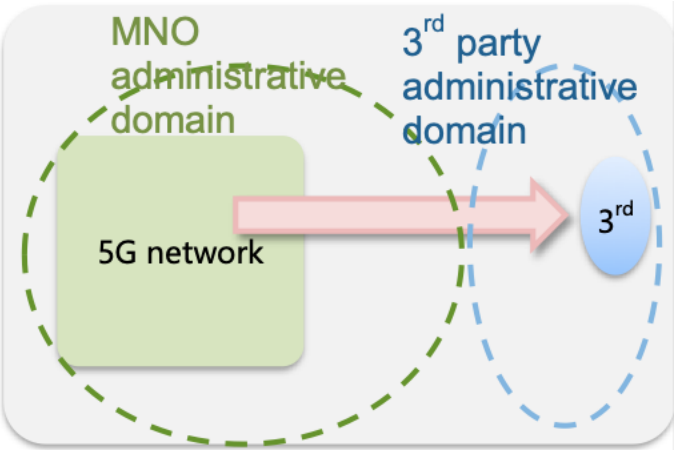


Figure 8. Level 1: Read: Can passively access exposed network service and functions (Golic, 2018).

Level 2: Semi-active exposure – The 3rd party can customize and accordingly change, provision or manage the configuration parameters of exposed elements of the network service and functions or network management services and functions. It is also allowed to have access to some exposed network management capabilities to accordingly customize, provision and update the configuration parameters of the network service and functions. This can relate to a closed part of the edge network. Semi-active access to exposure can be implemented by a NEF interacting with NFs via APIs or slice management interface. (Golic 2018.) Figure 9 presents how administrative domains start to overlap.

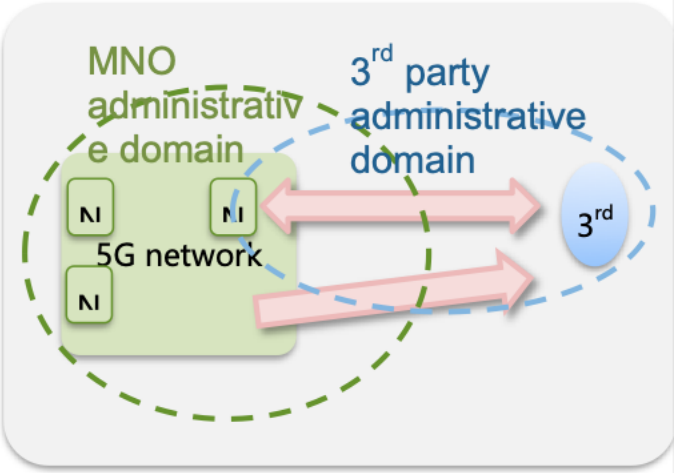


Figure 9. Level 2: Read/write/manage. In addition to Level 1, can configure and manage capability exposure and can access the network management capabilities (Golic, 2018).

Level 3: Fully-active exposure – The 3rd party is allowed to add, install and manage network access and communications services and functions and network management services and functions, based on exposed network capabilities, e.g., the exposed (hosting) network infrastructure. This can relate to a case where 5G core network supports connections to Local Area Data Network (LADN). In this case, MEC services are available only to mobile devices that are in the specified area. (Kekki et.al. 2018.) It is called as MEC/LADN scenario. It is preferable that the 3rd party should not be allowed to add, install and manage new elements of the 5G core network, especially critical/sensitive network functions. (Golic 2018.) Figure 10 illustrates how administrative domains overlap, when the 3rd party service provider is allowed to use MNO's resources.

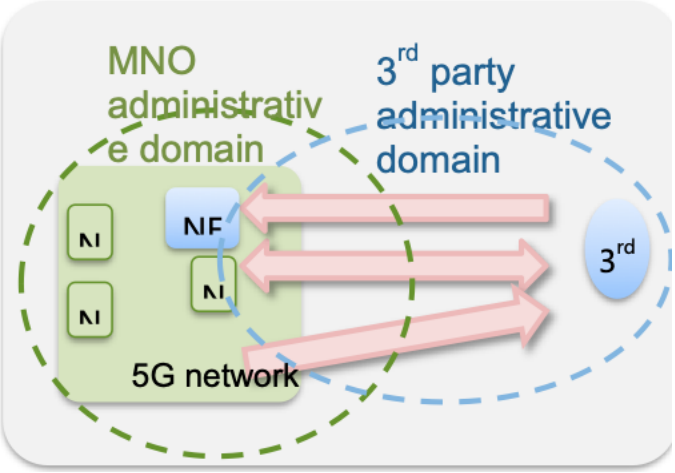


Figure 10. Level 3: Read/write/manage/provide. In addition to Level 2, can add and install network services and functions, and network management services and functions (Golic 2018).

4 SECURITY THREATS IN 5G CORE SLICING, MEC AND EXPOSURE

It is nothing new in the IT business that there are interconnected networks and IT systems communicating with each other's over APIs. Many business sectors have opened their APIs to integrate to partner's systems to provide more and better services to end customers. 5G mobile network can be seen as a digital platform that is essential part of the modern service-based economy. Digital platforms generally enable creation of new kinds of services and business models and the influence of digital platforms in modern service-based economy is remarkable. Facebook and social media, or Spotify and music, Netflix and movie are some examples, and more can be found in payments, healthcare, hospitality and ecommerce. (Lampi 2020, 23.) Mobile enabled digital platforms are needed.

Opened systems and provided services can be at quite different levels. It is obvious that, for example, a restaurant's food ordering system does not need to be as secure as e-banking. Possible havoc if system compromises is quite different in these examples, and there security considerations are at different level. Mobile operators and their systems are many times seen close to e-banking requirements. It will even be highlighted in 5G that will connect many aspects of society through the network ranging from critical infrastructures such as e-health, transportation, and electrical grid systems to user environments such

as smart homes and handheld devices. (Suomalainen et.al. 2020, 2.) MNO's decisions and actions have great influence how comfortable feeling end user has. From this point of view, the security of the new technological opportunities needs careful consideration.

Chapter 3 described how SBA is a new distributed architecture to provide 5G mobile core services. It increases complexity due to the decompiling of network services/functions. It will also easily reduce MNO's overall control and visibility, as Communication Service Providers (CSP), 3rd party apps and enterprise customers may have control for MEC instances in the MNO core. As said, IT systems integrations over APIs is nothing new, but opening mobile core network over APIs to 3rd parties is new for MNOs.

This chapter describes security threats, attack surfaces and vectors from theoretical point of view based on literature. The validity of threats depends on how much of the reference 5G SBA has been implemented in the MNO's network. Not all the upcoming features will be implemented at once, but step by step and in a phased way. This theoretical chapter does not yet consider the validity of the threats, but they are addressed purely from the theoretical point of view. In the discussion (Chapter 7) these theoretical threats are reflected against the real world 5G core as implemented in 2021. Most likely this reflection becomes outdated once technology evolves and further studies should be conducted again after some years.

ENISA has created a common taxonomy of threats in 5G security and introduced that in their 5G Threat Landscape report in 2020. It is a framework and intended to help communication among the different stakeholders in policymaking, regulation, product development, system implementation and operation. The taxonomy consists of nine categories. They include different kind of hostile activities, both technical and physical, but also unintended damages, outages, failures and disasters. Figure 11 present all nine categories in more detail. The taxonomy mind map introduced in the ENISA's report is used in the conclusion

chapter when answers to research questions are derived from the interview findings. (ENISA 2020.) The mind map is presented in appendix 3.

Figure 11. ENISA 5G threat taxonomy categories. (ENISA 2020, 125).

4.1 5GC slicing threats

Slicing threats in 5G core side can be summarized in three major categories: isolation breakout, resource abuse / DoS and unauthorized access. Different variations of them are covered in the next chapters

4.1.1 Controlling Inter-Network Slices Communications

The purpose of the slice is that it carries user plane data, and it has ingress/egress communication on the user plane. Control plane data is needed to create the slice and its characteristics. A possible threat is that there are undesired communications, which then might have the potential to disrupt the functioning of the slice. (Harel et.al. 2016.)

4.1.2 Impersonation attacks against Network Slice Manager or Host platforms within an operator network

A Network Slice Manager is responsible for dynamically creating and destroying instances of a network slice and provision them to physical host platforms (e.g. routers, switches, servers). These host platforms will be deployed across the operator network and possibly across separate and distant geographic locations. The Network Slice Manager, as well as the target host platforms, must not be trusted. The Network Slice Manager should somehow know that the host platform on which a network slice is to be run, is an operator authorized platform. Also, host platforms must know that the Network Slice Manager with which they are interacting is authorized by the operator. Otherwise, impersonation attacks against Network Slice Managers or host platform systems can have devastating consequences for operators since they expose the network and the services supported by that network to corruption, removal, disclosure and interruption threats described in the previous section. (Harel et.al. 2016.)

4.1.3 Impersonation attacks against a Network Slice instance within an operator network

A Network Slice Manager would need to support provisioning functionality for already deployed and running network slice instances. This is needed to add subscriptions to an already deployed network slice instance. The Network Slice Manager must somehow guarantee that the correct and authorized instance of a given network slice is being provisioned. If that cannot be guaranteed, impersonation attacks against a network slice instance may impact all services supported by that network slice instance and would allow corruption, removal, disclosure and interruption threats. (Harel et.al. 2016.)

4.1.4 Different security protocols or policies in different slices

Different slices that offer different services may have different performance constraints, and different security requirements. These are for example the following:

- The service in one slice may require extremely low latency, which constrains the security protocol, e.g. affecting key derivation on service setup, or key management on inter-cell handover.
- The service in one slice may require extremely long device battery life, which constrains the security protocol, e.g. how often re-authentication is performed.
- The service in one slice may be very privacy-sensitive, requiring unusually intensive security procedures, e.g. very frequent reallocation of temporary identities.
  (Harel et.al. 2016).

While security mechanisms and requirements may vary between slices, it need to consider how well those slices are isolated from each other. There is a danger that if someone can attack the “lower security slice”, they may also impact the “higher security slice”. Furthermore, if someone can attack a “lower security slice”, they might be able to impact the whole network. (Harel et.al. 2016.)

4.1.5 Denial of service to other slices

By exhausting resources in one slice, an attacker may exhaust resources common to multiple slices, and hence cause Denial of Service (DoS) or service degradation in other slices, too. Common resources can mean either hardware-level resources (memory, processing power, storage space) or network functions providing services to multiple slices e.g., a single home subscriber server (HSS) providing authentication vectors to mobility management entities (MME) in multiple slices. (Harel et.al. 2016.) It should be reminded that the same result may happen by accidental configuration change if control methods are inadequate.

Denial of service may also have other objectives than just traditional denial. An attacker may want to “do something bad” in slice A. Normally, slice A would run its normal security protocols, and this would prevent the attack, but if the attacker can exhaust resources in slice B, in a targeted (and perhaps carefully timed) way, with the result that slice A is short of resource and unable to run its normal security protocol; now, perhaps, the attack in slice A can succeed. (Harel et.al. 2016.)

4.1.6 Side channel attacks across slices

Side channel attacks are a class of attack on implementations of cryptography. They occur when an attacker can learn something about cryptographic secrets by observing or influencing the platform on which the crypto code is running. This threat is possible when slices A and B share some underlying hardware. If an attacker can observe or influence how code runs in functions in slice A, he may be able to affect the running of code in functions in the slice B machine, or extract information about the running of code in slice B. This may allow side channel attacks – in particular, timing attacks – that extract information about cryptographic keys or other secrets in slice B. (Harel et.al. 2016.)

4.1.7 Sealing between slices when the UE is attached to several slices

User equipment (UE) security, while UE is attached to two or more slices simultaneous is not exactly question of the 5G core, because a UE exists at the RAN layer. However, this threat needs to be addressed here, because UE easily have an affect over core security as well. The UE could be attached to several slices, which may have various level of sensitiveness. If there is no separation in the UE between data communicated via different slices to and from the UE, then the value of separating slices on the network side could be reduced. For example, a UE may receive sensitive data via one slice and then publish that data via another slice. The situation is similar when a laptop has an Internet access on the one hand, and on the other hand it has access to an enterprise network. A route can be formed from the Internet to the company internal network via the laptop. As the UE has most likely no notion of slicing, then the policing of data inside the UE should be based on some other mechanism. (Harel et.al. 2016.)

4.2 MEC threats

While there are many threats against the MEC itself and UE using the MEC, the purpose of this thesis is specially to study what threats are valid for MNO. Those are billing risks, MEC resource exhausting, influence on the radio and break in/out to operator core.

4.2.1 Billing Risks from MEC deployments

In conventional cellular networks, billable traffic is routed into the core network. During roaming, it is usually routed into the cores of both the visited and home network. This allows both networks to keep track of how much data is being consumed and prevent billing errors, or fraud. With MEC, data is expected to be routed directly between the UE and the network edge, without passing through the core network (and without touching the home network at all in a roaming scenario). The visited network must rely on edge components to tell it what charging records to send to the home network, and the home network must also rely entirely on these components, despite having almost no control over how they are set up/secured. Since the edges of networks are more vulnerable to attack than the cores, this creates a significant risk both of billing errors and billing fraud. (Harel et.al. 2016.)

Under-billing is a risk where the end user or MEC application tries to use more data, or more valuable classes of data, than they will be billed for. Over-billing is another risk, for example if a hosted MEC application has a revenue-share model or pay-per-click model, it may try to inflate the amount of data billed for. Inter-operator roaming fraud may also be an issue. (Harel et.al. 2016.)

4.2.2 Third party applications on the same platform as network functions

Ultimate cost efficiency may lead to MEC model where edge computing applications will run on the same physical platforms as some network function. These will be third-party applications, not controlled by the MNO directly. There are risks of these applications exhausting resources that are needed by the network function. There are also risks of poorly designed applications allowing hackers to infiltrate the platform and hence affect the network function running on the platform, or even of malicious applications doing the same themselves. (Harel et.al. 2016.)

4.2.3 User Plane Attacks in a Mobile Edge Computing Environment

While MEC provides computation resources, it can also act as a content cache. The content or portions of it would move MNO's caches closer to the edge of the network. To minimize delays and jitters, it is likely that the current functionality of DNS resolution and content delivery networks would also move closer to the edge and therefore IP connectivity layer moves also closer to the user. In this new architecture, IP connectivity would terminate at the edge of the MNO network. This situation will alleviate challenges faced when optimizing encrypted video content end to end (UE to video server), since the content would now be delivered from replicas in the operator network. That said, a new set of challenges arise for the operator: security threats that target the content server using protocols like HTTP/HTTPS and security threats against content caches, i.e. cache poisoning attacks. (Harel et.al. 2016.)

Traditional attacks against caches were e.g. via HTTP response splitting, but now other type of attacks will be possible too. With MEC, a large number of caches at the edge of the network would be deployed, but most likely a single cache has considerable smaller capacity than used to be earlier, when there were few large CDN caches to serve large number of users. Smaller caches are easily to be overwhelmed by attackers with request for content not likely to be used by regular users. This would result in filling local caches with useless and unusable content for subscribers and would have the effect of disabling these caches. Such attacks can cause major disruption to the latencies and possibly committed SLAs. (Harel et.al. 2016.)

4.2.4 Sensitive Security Assets at the Edge

At the general level, compromised security assets at the edge may lead to different kind of spoofing, eavesdropping or data manipulation attacks. The attack surface increases the more functions there are as well as hosting computers. The same security risks also apply when sensitive security assets are exchanged between the core and the edge. (Harel et.al. 2016.)

4.2.5 Communication between the core and edge

As the physical connection from the core to the edge is outside of the premises of the MNO, the physical or virtual links are open to compromise. MEC orchestrator in the core to the mobile edge can be compromised if Trust Establishment between them is insufficient. (Harel et.al. 2016.)

4.2.6 Lawful Intercept requirements for MEC deployments

Operators are required to provide law enforcement agencies (LEA) support including Lawful Interception (LI) and retained data capabilities for traffic carried on their networks; typically this functionality is supported at nodes within the core network. (Council of the European Union 2019).

MEC will allow mobile phone networks to store and process contents in decentralized clouds in the vicinity of network users which can directly communicate with each other. Information will not necessarily be directed via central nodes, where lawful interception is currently implemented and hence would avoid the usual intercept points (Harel et.al. 2016). To overcome this issue, more LI points can be deployed, but placing multiple additional LI points around the network edge raises security risks. The more LI points there are, the more there are interfaces to penetrate to the network. If the MEC traffic must be available for law enforcement authorities, LI points need to be deployed at the edge nodes, which are likely to be more exposed to attack than core nodes (Council of the European Union 2019).

4.3 Threats in network exposure

NEF is like a gate to MNO’s network, its services and resources, so NEF’s security and protection is essential for the MNO’s network security. While poor authentication and authorization mechanism is an obvious threat (ENISA 2020, 142), a lack of the rate limiting is also a threat. Without a proper limit of the number of simultaneous active sessions and bandwidth, NEF API could be overloaded and therefore affected by DoS attack. As a consequence of DoS,

NEF API function can become unavailable, or some sort of buffer overflow may occur and whole authentication mechanism can collapse in the worst case.

The network capabilities exposure should be seen as part of a commercial contract. It needs to be clearly stated in the contract what is under control of the MNO and what can be exposed to the 3rd party. The MNO should maintain control of new network management functions and services provisioned by the 3rd party. The MNO must have the capability to control, in real time, that the exposure of API is compliant to the technical terms and conditions of the contract. (Golic et.al. 2018.)

In addition of above NEF specific threats, ENISA reminds 61 possible vulnerabilities not only for NEF, but also covering many 5G core functions, thus it is providing a good checklist for an operator. Issues in the list are not tricks how to penetrate to 5G core via NEF, but more like best practices what kind of threats should be considered and mitigated. Many of them are best practices type of items, like shut down of unnecessary network services or the implementation of the security event logging. (ENISA 2020, 143-163.)

5 SECURITY MECHANISMS AND PROCESSES

Transformation from legacy and proprietary telecommunication protocols towards IT industry standards, like HTTP/2, TLS, TCP and RESTful will most likely expand the potential pool of attackers, because protocols are wider used in the IT industry and better known. However, 5G has designed-in security controls to address many of the threats faced in earlier generation mobile networks. In the 5G core network this means that there is no assumption of safe and secured networks, but they all are considered to be open and insecure and all links could be tapped. “Secure by Design” principle in 5GC leads to use of mutual authentication. It means that the sender and the receiver can trust each other’s, and the end-to-end connection is encrypted. (GSMA 2021.)

5.1 Evolution of the trust model

As mentioned in the previous chapter, 5G has secure by design and that can be understood from the 5G onion model of trust. In this context it is important to present radio parts as well, even though they are not in the scope of this master's thesis. Trust within the network is considered decreasing the further one moves from the core. (3GPP 5G Security 2018.) The onion model is presented in Figure 12.

The trust model in the UE is simple: there are two trust domains, the tamper proof universal integrated circuit card (UICC) on which the Universal Subscriber Identity Module (USIM) resides as a trust anchor and the Mobile Equipment (ME). The ME and the USIM together form the User Equipment (UE). (3GPP 5G Security 2018.)

The Radio Access Network (RAN) is separated into distributed units (DU) and central units (CU). DU and CU together form the 5G base-station (gNB). The DU does not have any access to customer communications as it may be deployed in unsupervised sites. The CU and Non-3GPP Inter Working Function (N3IWF – not shown in Figure 12), which terminates the Access Stratum (AS) security, will be deployed in sites with more restricted access. (3GPP 5G Security 2018.)

In the core network the Access Management Function (AMF) serves as termination point for Non-Access Stratum (NAS) security. Currently the AMF is collocated with the SEcurity Anchor Function (SEAF) that holds the root key (known as anchor key) for the visited network. The security architecture is defined in a future proof fashion, as it allows separation of the security anchor from the mobility function that could be possible in a future evolution of the system architecture. (3GPP 5G Security 2018.)

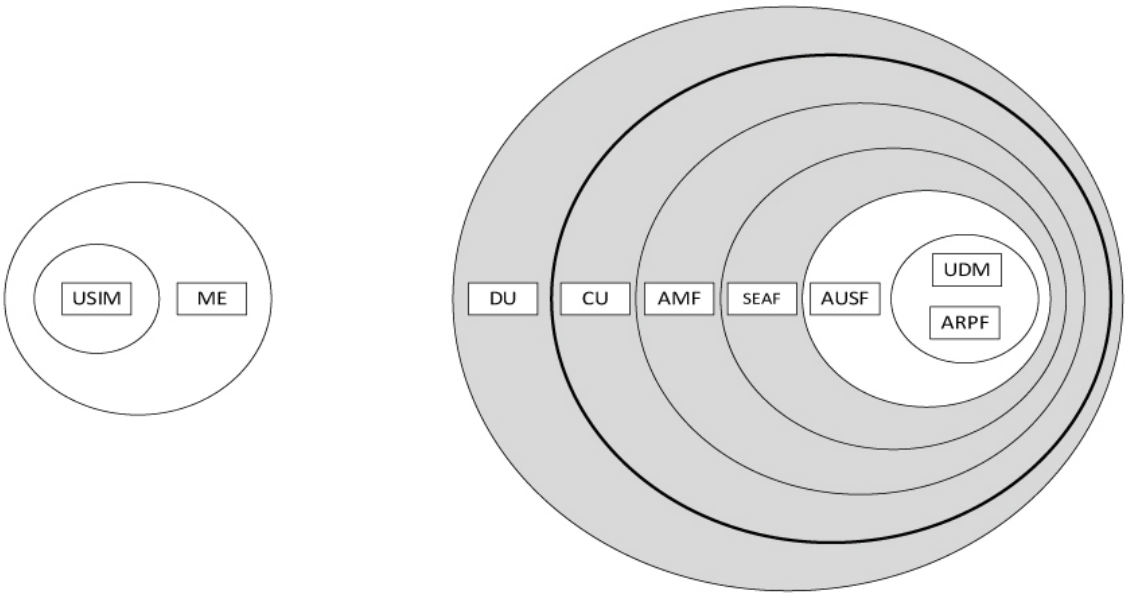


Figure 12. Trust model of non-roaming scenario (3GPP 5G Security 2018).

The AUthentication Function (AUSF) keeps a key for reuse. Authentication credential Repository and Processing Function (ARPF) keeps the authentication credentials. This is mirrored by the USIM on the side of the client, i.e. the UE side. The subscriber information is stored in the Unified Data Repository (UDR). The Unified Data Management (UDM) uses the subscription data stored in UDR and implements the application logic to perform various functionalities such as authentication credential generation, user identification, service and session continuity. (3GPP 5G Security 2018).

5.2 Protecting network functions (NF)

The Network Repository Function (NRF) is a key service of 5G networks. It is responsible for registering new network functions (NF) and storing their profiles. If TLS certificates are not used and NRF does not perform service authorization, it is possible to impersonate attacker’s NF to the core network. This, of course, also requires from the attacker to find a way to connect to operator’s core, but when MEC requires stretching the operator’s core to customer premises, this is a real threat to be considered. (Positive Technologies 2020, 10.)

If the attacker can impersonate to the operator core and no TLS and authorization used, a data leak may be caused, by obtaining NF profiles or even service disruption by deleting NF profile. (Positive Technologies 2020, 11).

Prevention of the previous scenario should be easy. Network functions should always be required to authorize services by verifying transport layer security (TLS) certificates when a connection is being established. When NRF is using authorization, other network services must verify sender service identity when they are receiving incoming requests. (Positive Technologies 2020, 11.)

5.3 Network slice lifecycle

In the very early stage of 5G slicing, it is possible provision slices manually, because there are only a few basic slices. However, the target must be that on one day slicing is a common service in the operator's network and new slices come and go. Then the slice lifecycle matters. Figure 13 presents 3GPP's overview how slice lifecycle proceeds.

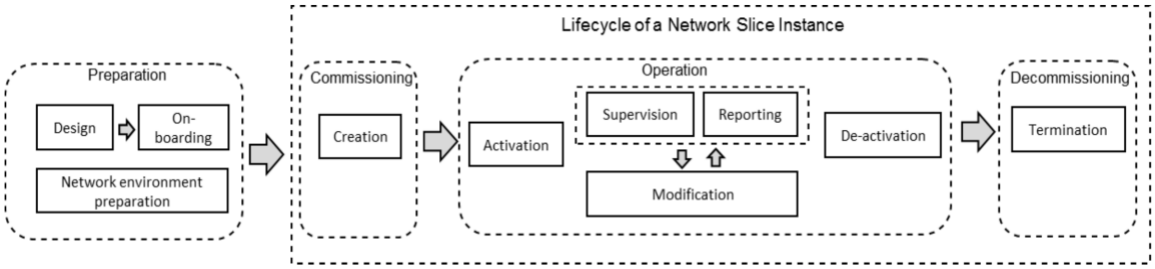


Figure 13. Management aspects of network slice instance (Tovinger et al. 2018).

In the preparation phase the network slice instance does not exist. The preparation phase includes network slice template design, network slice capacity planning, on-boarding and evaluation of the network slice requirements, preparing the network environment and other necessary preparations required to be done before the creation of a network slice instance. (Tovinger et.al. 2018, 3.) Threats to be targeted at this phase are malicious images and templates (Kler 2020).

Provisioning in the commissioning phase includes creation of the network slice instance. During network slice instance creation all needed resources are allocated and configured to satisfy the network slice requirements. The creation of a network slice instance can include creation and/or modification of the network slice instance constituents. (Tovinger et.al. 2018, 4.) Threat to be targeted at this phase is misconfiguration, which can lead to weakening of the network overall security (Kler 2020).

Operation includes the activation, supervision, performance reporting, resource capacity planning, modification, and de-activation of a network slice instance. Provisioning in the operation phase involves activation, modification and de-activation of a network slice instance. (Tovinger et.al. 2018, 4.)

The decommissioning phase includes decommissioning of non-shared constituents if required and removing the network slice instance specific configuration from the shared constituents. After the decommissioning phase, the network slice instance is terminated and does not exist anymore. (Tovinger et.al. 2018,4). Threats to be targeted at this phase are associated with data management: how sensitive data removal is ensured, when a new slice is provisioned on resources previously used by other businesses, and what happens with logs / monitoring data. (Kler 2020.)

If network slices require isolation from each other’s, then breakout is a security threat to be managed. The normal solution is that user equipment (UE) is allocated different identities according to the slice type and differentiator. The differentiator defines if different slice types are allowed to be connected to the network at the same time or allowed to have active data connection at the same time. This also leads to the best practice that slices that have very different levels of sensitivity should not be co-hosted on the same hardware platform to avoid side-channel attacks. (5G Americas 2019.)

5.4 Protecting the mobile edge computing (MEC)

A large scale of possible use cases and deployment models make it challenging to describe accurately different security mechanisms to protect MEC system. Key challenges are that MEC is located at the weaker physical security locations than in the past at the operator's data center. The movement of user-plane (UP) functions to the network edge pushes sensitive data to the edge too and therefore increasing the risk for sensitive data to compromise. Figure 14 shows how the 5G UP is pushed close to the edge. This makes cross-layer attacks possible against MEC platform, applications, user plane functions (UPF) and 3rd party applications. Because of those, MEC needs multi-layered defense in depth approach. Different security mechanisms are needed at all layers: network, MEC platform, multi-tenant operation and management (O&M), user plane function (UPF) and application. (Kler 2020.) Data from privacy-critical applications should be stored and processed in MEC servers, which are within trust zones. In untrusted zones should be run only applications with most critical latencies. If data backups are necessary in MECs in the untrusted zone, they should be stored and processed within cloud in encrypted form. (Suomalainen et.al. 2020.)

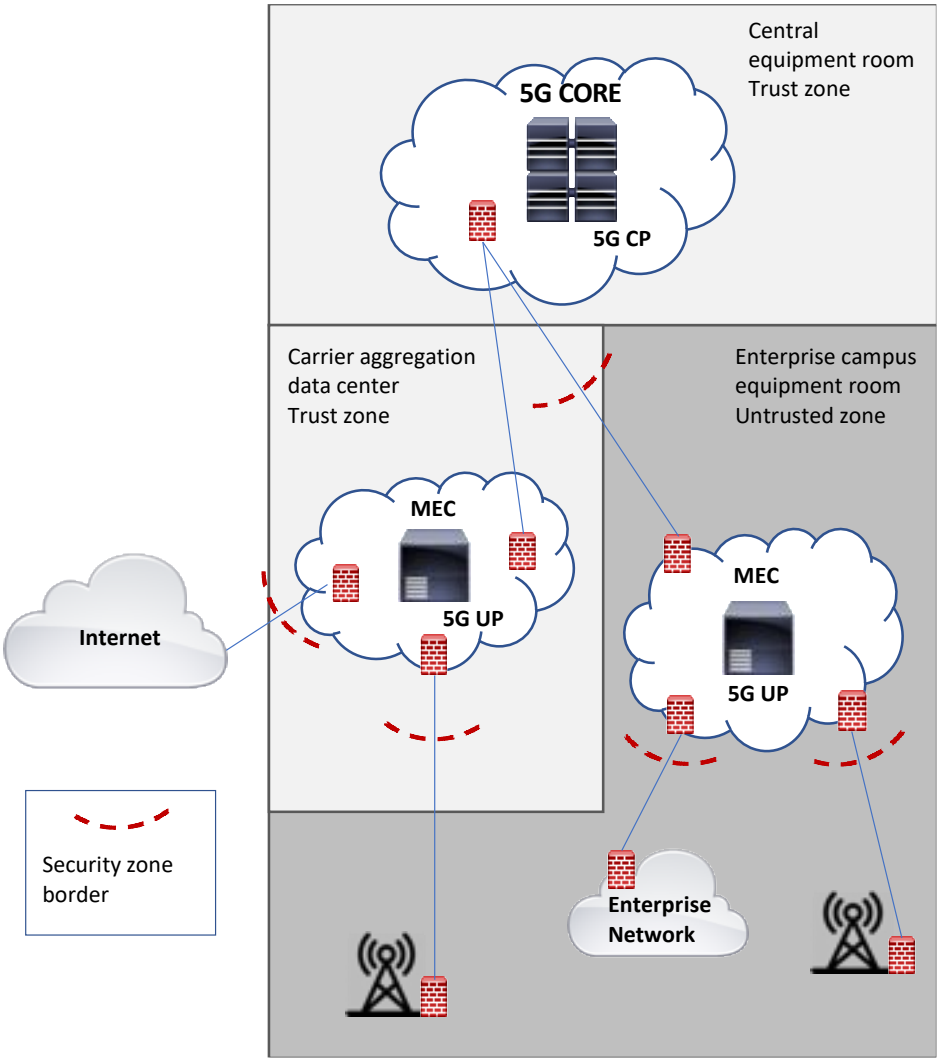


Figure 14. MEC trust zones and how to protect mobile edge (Kler 2020).

Multiple overlapping security mechanisms should be implemented to ensure MEC security. A hardware isolation deployment is recommended with separate security zones. An efficient way to accomplish multiple security zones are virtual firewalls. Security zone isolation reduces distribution denial of service (DDoS) attack surface and limits possible damages in a smaller area. It also creates a strict separation between MEC platform network, O&M network, user plane function (UPF) network and external data networks like customer's own data network or Internet. Firewalls should also deny Internet access if MEC application does not require it. Virtual firewalls not only limit allowed traffic, but also create control points to collect log information and detect if something malicious is going to happen. Logs itself are not enough, but they need to be submitted to Security Information and Event Management (SIEM) system. (Kler 2020.)

On the MEC host there should be implemented software signature verification and only signed software is allowed to run. It ensures that no malicious software can be implemented in the MEC. The software running on the MEC has some APIs to communicate with mobile end devices. API access authorization should be implemented as well as API flow control. Once these are created, KPI monitoring should be implemented to detect and avoid malicious resource occupation. This information should be submitted to SIEM system as well. (Kler 2020.)

5.5 Machine learning (ML) based security solutions

Mobile networks are becoming more complex all the time and equipment number is expanding, both network devices and end-user-devices. This leads to the requirement of the automation in the network management. Machine learning (ML) is needed to make effective automation. ML is expected to mitigate human-control risks, and empower mobile networks to self-control, adapt, and heal themselves with changing user, service and traffic requirements. (Suomalainen et.al. 2020.)

The feasibility of ML depends on the quality of data. In complex mobile network, collecting realistic and comprehensive data sets is often a challenge. ML also introduces major maintenance challenges in complex settings. Data sources may become unstable over time and have dependencies that are difficult to analyze. Similarly, models and ML-based systems may be entangled, and small changes may lead to unexpected situations and vulnerabilities. ML is by its very nature statistical, predictions are always possibilities, and in the case of many varieties of learning algorithms, the amount of error is unknown for new data. If the underlying causality of ML remains obscure, it is possible that output may not reflect the intended cause but may be something completely different with an accidental correlation with it. This kind of fault is difficult to detect since the model might still yield good results. (Suomalainen et.al. 2020.)

Training is essential that ML would be effective. There are several techniques to train ML, but for example in adversarial training, malicious samples are included in the training data. The approach requires that defenders can collect or generate valid examples of known attacks. Sophisticated and targeted attacks inside 5G networks have been quite rare. A challenge may then be how to acquire or generate realistic adversarial examples. Therefore, it would be important that operators cooperate and share information on detected threats and adversarial samples. Also, honey-pot techniques can provide an approach to collect adversarial samples. (Suomalainen et.al. 2020.)

Timing, when, how wide and what kind of ML system to utilize, is a good question. Inherent protection for 5G networks comes from its partially closed nature. Network components, interfaces, and functions – including ML software – are not available for everybody. 5G networks incorporate various platform and communication security solutions protecting the integrity of the platform and data and for keeping external adversaries outside. However, the size and complexity of 5G networks have left the networks partially open to advanced adversaries. Persistent adversaries will eventually find weaknesses in the large attack surface of 5G. For example, nation-level agencies have the same capabilities as the defenders and may, e.g., purchase or otherwise acquire the same ML software that the defenders are using and use it for stealth testing and rehearsing attacks. Consequently, a single layer of defense is not likely to be sufficient. Vertical 5G perimeter defenses must be enforced with ML-based security applications that protect ML functions and 5G platforms from threats coming from inside, as well as with approaches for robustness and resiliency of ML algorithms. (Suomalainen et.al. 2020.) Suomalainen et.al. summarize in their report 2020 that even though ML has great capabilities and it will be definitely needed, it has still open questions and further research is needed how to overcome limitations.

5.6 Situational awareness

The 5G network security should be based on domain specific security thinking. It means that there are different zones with different security levels. Generally, domains are radio access network (RAN), 5G core (5GC), MEC and slicing,

which goes through RAN, 5GC and MEC. Figure 15 shows how security zones can be defined. It divides 5G network into four different security zones and each of them should be monitored by Security Information and Event Management system SIEM.

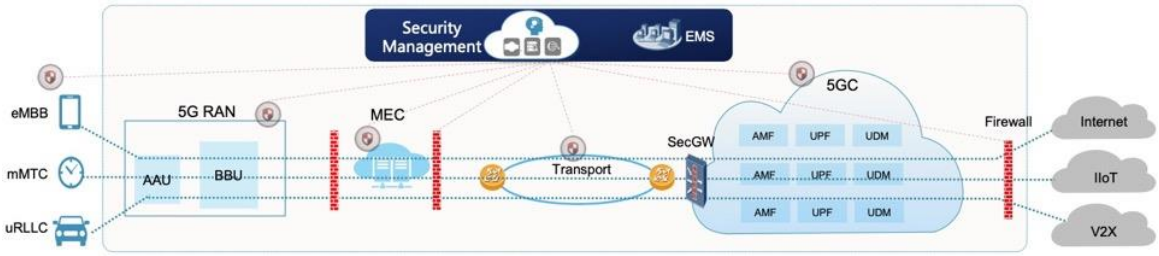


Figure 15. 5G multi-layered security zones (Kler 2020).

Security Information and Event Management (SIEM) is nowadays the most used system in security operations centers (SOC). It collects event and log data created basically any kind of IT system that can create log information. SIEM is used to collect all the data in one place, analyze it and raise alarms at the right priority according to the issue severity (Muikku 2020).

Efficient situational awareness overview can be constructed when SIEM information is enriched with artificial intelligent (AI) and/or machine learning (ML) capabilities. 5G network will be a complex system and that is why Kler recommends MNO to form 5G threat overall awareness. It contains an overview of all network elements and visualize suspicious or malicious occurrence and trends. A network-wide threat awareness requires that visualization covers security awareness of all elements: radio domain, transport, core network functions, signaling and service plane, and MEC environments. Figure 16 illustrates a conceptual overview, what kind of key point interests (KPI) would be useful to be presented in a situational awareness system screen.

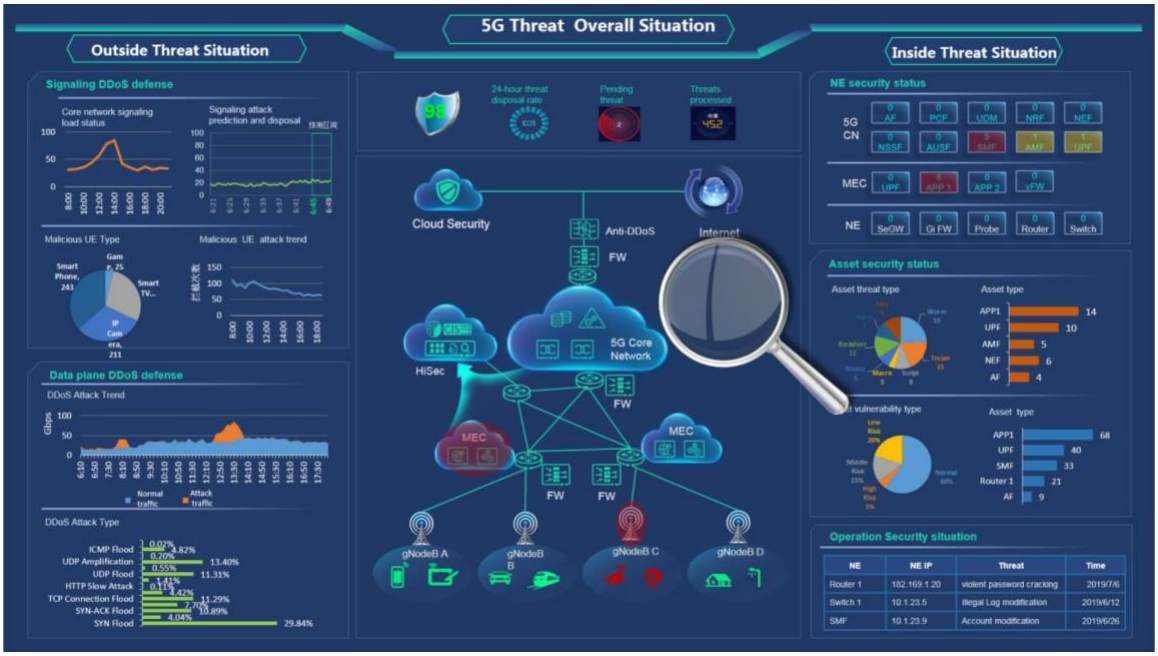


Figure 16. Simplified principal example of the 5G situational awareness system overview (Kler 2020).

6 DATA COLLECTION

In the first half of the 2021 5G stand-alone core networks are still under construction and only first steps towards production networks have been taken. The full scale 5G SBA model in the production and all the services it can offer are the result of an evolution path and is expected to be completed in the coming few years, as pointed out by Kalle Lehtinen, CTO at Elisa (Kokkonen 2020). This means that best data collection method of upcoming architecture, implementation phases and security mechanisms are theme interviews of the relevant technology area experts.

In the semi-structured interview, there are no strict structured questions to answer, but a schema that the interviewer uses. Theme interview can also be considered a discussion which is led by the interviewer, but the knowledge comes from the interviewee. The objective of the interviews is to gain knowledge that experts have. (Alastalo et.al. 2017.)

It should be defined, who can be considered as an expert, but there is no exact answer for that. It is a status that can be achieved from science, professional or institutional roles (Alastalo et.al. 2017). According to Alastalo et.al. it is important to find out who has such knowledge or experience that can be considered as a professional. In this thesis, professionals were selected according to their job roles, companies they work for and recommendations by other professionals. It was an intension that interviewees are 5G core network experts from different angels of the topic, rather than everyone to be a security expert. That ensures that there is different point of views to fulfil interview targets. The first target was to obtain a holistic overview based on professionals' experience and method was systematizing interview (Alastalo et.al. 2017). The second target was that interviewees share their opinions in general, certain security threats and possible counter measures. That required that the interviewer was well prepared and was considered as a professional as well, instead to be just a student (Alastalo et.al. 2017).

Due to the commissioner of the researcher, it was easy to find most experienced professionals from one MNO and mobile technology vendor as well. For the same reason, professionals from other MNOs were not able to get interviewed. All the interviewees have 20 years or more working experience at the telecommunication industry and two to ten years in the relevant expertise area. The 5G is the most interested technology in this research, but because it has been available so short time, the 4G technology expertise was required as well. Due the COVID-19 pandemic all interviews were conducted over Microsoft Teams and they were recorded. Teams is actually a good tool, video call provides facial expressions almost like in live interviews and they are also recorder opposed to live interview, which normally are just voice recorded, but not video recorded. A traditional phone interview suffers lack of facial expressions, which is many times mentioned as a lack of remote (phone) interviews (Ikonen 2017). Teams also reduced time frame needed to carry out interviews. Interviewees 1-4 live in Finland, but not all in the same city and interviewee 5 lives in the other country. Interviews were carried out in May 2021.

Identities of the interviewees are not included in the report, but their roles and later references are:

Intv1: 5G core and MEC lead architect
Intv2: Mobile core tech lead
Intv3: Mobile core security manager
Intv4: Mobile core engineer, security
Intv5: Security architect, network slicing & MANO

The data collection method was selected according to the research objective. This report is to study 5G core network security risks in the network slicing and MEC cases, and how them could be provisioned over NEF. Therefore, the way how interview questions are answered is not relevant. The capture of nonverbal elements and context of the interview itself are considered out of scope for this study. Instead, the facts and point of views said by the interviewees are considered interesting. (Alastalo et.al. 2017, 181-197.) All the interviewees participated willingly and were motivated to provide all their knowledge of the topics covered during the interview.

Semi-structured thematic expert interviews were the data source for this master's thesis. Interview themes were derived from the literature review that was completed before interviews. The defined themes were 5G SBA model, 5GC Slicing, MEC, NEF and security aspects of all of them. The semi-structured model worked well, because all the interviewees had partly different experience and expertise, so strictly structured questions would not work. The semi-structured model provided flexibility to approach themes from each interviewee's own expertise point of view and they could provide all their knowledge to the researcher. The researcher owns over 20 years' professional experience from telecom industry and that also helped to make further questions from the answers and determine when there would be no more relevant questions for a certain topic. Interviews were carried out in Finnish or English, depending on the interviewee's native language. Interviews started with a warmup type of background questions and then proceeded into the actual themes. All the interviewees had good and deep knowledge of the 5G SBA model, but depending

on each interviewees' job roles and expertise, on the next topics emphasis was more on slicing, MEC or security. The target length of the interview was around 60 minutes. It was not a strict limit and actual interviews were between 48 minutes to 75 minutes. Each interview meeting also contained a short introduction of the researcher and his master's thesis studies as well as the master thesis research question and methods. That part took 5-10 minutes and not included in the recorded interviews.

The number of interviewees was quite low for a reason. The researcher wanted to interview only true professionals and keeping that in mind, increasing the number of interviewees would easily cause that someone with less experience would have been interviewed. A good question would be what the benefit of the greater number of interviewees would then be. Traditionally, a large number of interviewees is needed because grounded theory-based approach had required a high saturation and had been adopted as a general guide for other types of research interviews as well (Hyvärinen et.al. 2017). However, Hyvärinen also challenges that traditional approach and mention that even just one interview would be enough and well defined and cropped research problem also helps the research identify when saturation can be obtained with s small number of interviewees. Large number of interviewees does not automatically mean better results. Basing of those theses, the researcher decided to stay this number of interviewees.

6.1 Interview semi-structured theme questions

Theme questions were formulated prior to the interviews to maintain discussion and lead the interview sessions through topics that are relevant from the research questions point of views. Questions were not asked by strict pre-written sentences, but interviewees were rather encouraged to tell what they consider to be essential from the following topics. Even though the researcher heard something very interesting or brilliant from some interviewee, the researcher did not steer another interviewee to comment on the same thing. By this decision the researcher aimed to obtain as many different aspects of the following themes as

possible and ensure that interviews do not repeat each other. The themes of the interviews are listed below.

1. **Background information**
   - educational level
   - working experience in general in the industry
   - working experience with the relevant technologies
2. **5G SBA**
   - SBA readiness
   - what parts are ready to be implemented
   - is/can/should SBA implementation to be phased
   - does it need to be implemented everything at once
   - multi-vendor interoperability
   - management
3. **5GC Slicing**
   - purpose of the core slicing
   - management methods
   - automation levels, provisioning
4. **MEC**
   - purpose of the MEC
   - deployment models
   - MEC equipment
   - provisioning
   - management
5. **Security considerations of slicing, MEC and NEF**
   - NEF as 3rd party interface
   - how to keep MEC secured, if 3rd party can utilize their own code in there
   - protocol evolution from traditional telco protocols to ip-based protocols, what it all means?
   - AI- and ML-based security mechanisms, nowadays or future?

6.2 Data analysis

The data analysis for qualitative data may vary drastically depending on data and branch of science. This thesis researcher trusts this citation when considered selected methods to be justified: “Unlike quantitative analysis, there are no clear rules or procedures for qualitative data analysis, but many different possible approaches” (Spencer et.al. 2014, 270). Qualitative data analysis can also have many aims, such as to describe a phenomenon in greater detail, compare several cases, explain a phenomenon, or develop a theory of a phenomenon (Kohn et.al. 2018, c. 4.2). In this research the data analysis aims to explain 5G core slicing, MEC and NEF, which are the phenomena, and develop answers to what does it means for MNO’s security, i.e. to develop a theory of a phenomenon.

In many qualitative research cases transcriptions are created from interviews before further analysis. It is especially used in social and human sciences when focus is on the expression or context of the interview situation. However, this research concentrates on facts, words and the way how the expert tell something is irrelevant. Conclusions directly from the interview recordings are possible when the number of interviewees is limited, as in this thesis. (Hirsjärvi et.al. 2015, 138.)

Interviews were conducted in two weeks, and the first listening of the recordings were in one week after the last interview. During the first listening, key points of each interviewee speech were noted on Excel spread sheet, including a time stamp when interviewee said that. The first review was done quickly after interviews, when interviews were fresh in the researcher's memory. Later recordings were listened completely two more times in the next four weeks, notes were detailed to deeper level and themes created. When writing theme sections 6.3 – 6.6, recordings were listened to at specific time stamp to verify the exact statement of the interviewees.

The common way is that themes, categories and coding are emerged from the data during the analysis phase, but also a framework method exists. In the framework analysis pre-adopted concepts can be assigned from the literature and there are predefined categories to which the data are coded. (Kohn et.al. 2018, c. 4.3.3.) The following theme-sections were predefined from the research questions and the literature review and are compiled according to interviewees' statements. The approach is deductive.

Answers to one question might have been long and wide and many times provided statements and aspects to multiple themes. During listening and compiling, interviewees' comments were compared to the literature review or other sources. The researcher decided to use this kind of method, because it gave an opportunity to evaluate interviewees' statements if they make sense and are accurate or if there occurs deviation from the reliable sources or between interviewees. This is also a base for data reliability assessment. No incredible

statements occurred. The researcher considers interviews credibility high because interviewees were very open to say if they had a feeling that they did not know much about some of the interview topics. That helped to use more time on the areas which each interviewee knows the best. In addition, that ensured to achieve the goal of the expert interview: gain the knowledge that expert have.

6.3 Theme 1: SBA Readiness

The 5G core network is based on SBA and cannot be implemented in any other ways. All further functions like NEF and services like slicing or MEC require that SBA is implemented in the network. That is why it was essential to find out how interviewees saw SBA readiness, what can be implemented now and how it will be evolving in the future. Even though commercially 5G has been available since 2019, it has been based on the 4G core, i.e., so called non-stand-alone NSA-5G, and SBA based 5G core, stand-alone SA-5G is under construction and still commercially unlaunched in many MNOs (on summer 2021).

All interviewees stated that SBA is ready for 1st phase deployments both from standardization and product availability points of views. Rel-15 has been freezed since July 2019 (3GPP Release 15 2019), and vendors have released hardware and software to fulfill requirements. Intvs 2 and 4 reminded that even though the core side would be ready, there are still some stability issues with certain mobile devices, when network is in SA-mode. Intv4 pointed out that vendors are bringing new software in high cycle now and each of them are essential to eliminate issues. Each update needs to be pre-production tested before being implemented in the live network. This causes pressure for testing, because nowadays it should follow DevOps -methodology rather than previous generations quarter-based update cycles, reminded by intv3.

In theory SBA allows multi-vendor environment because interfaces are standardized and that was also understanding of all interviewees. From the technical operability point of view, every interviewee thought that in the first phase single-vendor environment would be safer, because there might be slightly different ways to implement some features. This might appear in a situation when

some network functions can interoperate and work correctly in single vendor, but in multi-vendor it requires that all protocols comply standard in the same way. Intv5 took an example that one vendor might have implemented some proprietary feature, because it is not available at the current rel-15 standard but is still under standardization and is coming in rel-17. This means more careful testing at the operator side, compared to single vendor, where operator can rely more on vendor's own testing. This all might have some consequences for timetables and when new services are available to customers, intv4 compared prevailing time to race. That is one reason for a single vendor. Nevertheless, Intv1 mentioned that because of prior mobile generations in the operator network and multi-vendors in there, even in the initial phase, SBA needs to be multi-vendor. Intv4 reminded that from the commercial contract point of view multi-vendor system used to be more cost effective and that should be the priority in the SBA as well.

SBA based 5G core management will change. intv3 described that in the past there were one service per appliance and management consist of appliance management and service configuration tasks. Now in the SBA everything is container based on top of virtualization platform. That brings additional layers and centralized locations to control and manage systems. This all means that security requirements will increase and those are covered in the section 6.6.

Currently, in the initial phase, 5G core management can be done manually, because once the SBA based core is up and running, it remains quite static. Also, micro services in containers can be handled manually now, for example scaling service capacity in / out is rarely needed. But anyway, all interviewees saw that automation is required in the future, only the time span varied from couple of months to two years. There are three different drivers for management automation, as listed below:

1) **Security**: management automation to ensure that certain security policys are applied before and after updates. Getting more and more important now when update cycles are getting faster. (Intv3)
   Management automation can detect if some network function is compromised, shut it down and replace it by new non-compromised instance. (Intv5)

2) **Network / service reliability**: 5G core is complex and automation reduces risk of human configuration errors. Another vision is that if (other than security reason) container needs to be shut down and re-created, automation can detect that and apply the task. (Intv2, Intv4)
3) **Decreased lead time**: automation will be basic requirement, when NF resources will be provided to 3rd parties, for example slicing customers. (Intv5)

SBA core monitoring differs from previous generations' clear text diameter and SS7 protocols. Now protocols are on top of TCP layer and encrypted. In the past there were separate devices attached with each other's over external physical connections and now containers are running in virtual hosting platform. This means that external TAPs and basic sniffer like Wireshark are useless, because there is no place to connect TAP and capturing encrypted data without encryption keys does not provide any useful data. Intv2 mention this to cause two consequences: a) monitoring system needs be capable to unencrypt SBA core traffic, b) all traffic may not be encrypted in the first place if point A is not fully working. Intv2 mention monitoring to require new kind of thinking, because by default, and for security as well, the traffic between NFs stays in the virtualization platform and virtual TAPs are the solution to collect traffic to monitoring. This means also that monitoring solution must be part of the mobile core solution and its load for virtualization platform must be considered.

6.4 Theme 2: 5G Core Slicing

5G Slicing is one of the new services with highest expectations. Virtualized 5G network can open many new possibilities for services and business opportunities. Benefits on the radio side are obvious once it will be possible to guarantee capacity over the radio network and many research focus on radio resource virtualization. (Guan et.al. 2018.) Benefits of the 5G core network slicing are not so clear and end-to-end (E2E) slicing is also a bit unclear (Guan et.al. 2018). It was asked from the experts, what kind of reasons they see for core slicing. Regardless of the drivers for core slicing, it was discussed, how core slicing should be done and what security aspects there are.

All interviewees confirmed that the current release of the SBA is capable to fulfill basic core slicing functionality. MNO should start with 3GPP's recommended three different slice types: eMBB, URLLC and mMTC. They are easy to create and maintain manually and gain experience of the core slicing for further services in the future.

To the core slicing, four reasons were introduced:

1) **E2E SLA** will be possible (together with radio slicing). This means that UPF is the most important function to slice in the core. (Intv1). Certain type of traffic from one customer can be routed or restricted in one way and other types of traffic from the same customer can be treated other way. (Intv2)
2) **HSS slicing** would be possible. It is almost like full own mobile network for the slicing customer, but without need for hardware infrastructure and maintain work. (Intv1)
3) **Local breakout** is rising new technic and slicing is one key element to provide that. (Intv1) Low latency demanding services are depending on core slicing and local breakout. (Intv 2)
4) **Security** improvements (both MNO and customers), traffic can be directed to different slices depending on the traffic type, device type or purchased service and DoS attacks may saturate only one slice, while others stay unharmed. (Intv4). Core slicing allows to restrict customer data only in certain parts of the MNO network and sliced network is like a VPN for the end customer. (Intv2, Intv5)

None of the interviewees could be sure, what would be the first real need for the core slicing. It was more like a discussion of possibilities and future will show how the technology adoption will proceed. Many of the possible use cases are close to lowering the latency, but they need to accomplish conjunction with the radio slicing and if it is after all enough for latency to slice radio only?

Intv1 and Intv2 described rising local breakout demand. That is the technique, where mobile data no longer travels long path as it used to be in 4G. The 4G mobile data path is presented in Figure the 17 and local breakout data path in Figure 18.

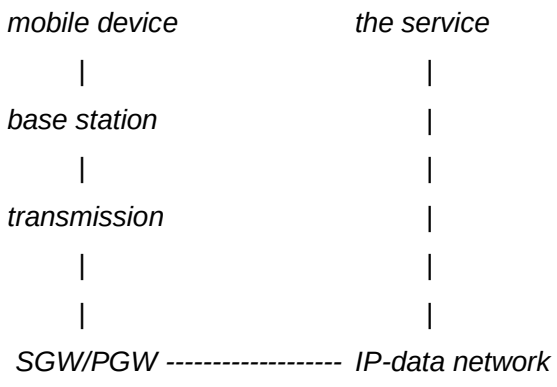


Figure 17: Mobile data path without local breakout. (Intv2 description)

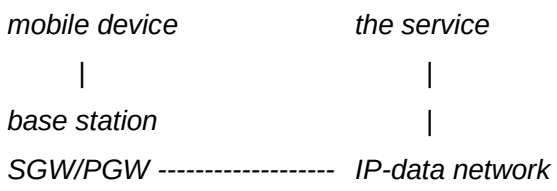


Figure 18: Mobile data path with local breakout. (Intv2 description)

Especially the transmission part can be a long haul for data from one part of the country to other side to SGW/PGW and then IP-data network takes the same distance but opposite direction. That causes propagational delay because in the worst-case data must make hundreds of kilometers round. The local breakout means that in 5G UPF its PGW-UP / SGW-UP can and will be located near the service and data path from mobile device to the service is significantly shorter.

End-to-end SLA and security are considered as different things, but in here they have much common. Intv1 and Intv2 saw that E2E SLA might be one of the strongest drivers for slicing and Intv4 raised improving security. When looking at the grounds, there are same mechanisms to produce both goals. Traffic classification and then different treatment is the basis for both. Certain traffic is driven to certain slice that has certain characteristics. Characteristics can define possible routes to ensure delays, but it is also a security feature, because data path can be restricted to certain parts of the network. Slice characteristics can also define single data flow maximum speed or customer’s total bandwidth. That is often to ensure that queuing and prioritization features can work, but as well it protects against DoS, while only one slice might be saturated, not whole physical data pipe or service.

6.5 Theme 3: MEC

MEC can be implemented with or without 5GC slicing, but most of times MEC is presented as one of the value-added services in the sliced network. Questions and discussion were structured to deal with applications and secure design principles.

All the interviewees mentioned the very same thing as main purpose for the MEC. It will be done to lower end-to-end latency between mobile device and service to as low as possible. Intv5 mentioned, when considering autonomous driving and the delay requirement is less than 5 ms, processing data in MEC is the only way to ensure such a low delay.

Offloading traffic from the central core closer to the edge has been ETSI's scenario for MEC around ten years, for example video stream at the stadium event to the crowd. However real implementations have been very rare. (Intv1) One of the reasons for this might be that ETSI MEC has its own APIs, while the mainstream is on Amazon Web Services (AWS), Azure or Google Cloud. All of them have a hybrid model, meaning that they offer edge server or cluster near users and that edge to be connected to public cloud over operator's edge routers. These big players are looking for MNO partners to get closer to mobile end users. Because mainstream cloud providers can also provide edge solutions, developers are more interested in those. It is easier for developers to use familiar APIs, this means also better cost efficiency for customers, when there are more choices over developers and they can work quicker, because no need to study less known APIs. The same complies also later life cycle during the solution maintenance. (Intv1).

From the MNO security point of view ETSI model MEC deployment means that MNO core network is stretched to the customer premises or at the mobile base station cabinet, which in many cases is not as secure as MNO data centers, where core devices have been traditionally located. MEC deployment is one more thing to be considered in MNO's physical security risk analysis: where MEC

is located at, is it in a locked cabinet, who can access it, what kind of physical access control there is? (Intv2, Intv3 and Intv4).

When considering cybersecurity of the MEC in general, there are two lines, corporate customers and consumer customers. As for corporate customers, security requirements are agreed with mutual contracts between agreement parties. This enables staying at a reasonable level and accepting certain risks, but of course the MNO must protect its own systems. However, if MNO is providing services to consumers, then it is general service and there are different security requirements for public telecom service, coming from Traficom. (Intv3).

Every interviewee underlined that MEC is still a future service without a clear use case today. There are ideas, for example video streams to be analyzed in MEC cloud and only relevant information to be transferred forward. That saves bandwidth and analyses can be accomplished quicker than if all information to be sent to centralized data center. (Intv2). However, the killer application that demands MEC and makes someone willing to pay for it, is still waiting to be found. Because there are lack of use cases, also questions what is provided by MEC or how automated the provisioning would or should be, are open questions. Technically everything is possible from pure memory, CPU and disk space to end user application, and manual case by case provisioning to fully automated from online ordering system. There are many possibilities and solutions depend on many aspects as described. (Intv1, Intv3 and Intv5).

6.6 Theme 4: 5G Core security, slicing, NEF, MEC

Security aspects were discussed under all themes during interviews, but outcomes are combined and analyzed in this chapter.

In the big picture 5G core is far more complex than 4G used to be and number of possible security risks is also greater. However, security thinking has also evolved, and lessons are learned from the previous generation. Security requirements are now divided into several sections because the architecture has changed so remarkably. The main partition is a requirement set for vendors and

another requirement for operators. That has been missing before. Security requirements for vendors are mainly coming from NESAS (Network Equipment Security Assurance Scheme, https://www.gsma.com/security/nesas-faqs/) , which is a voluntary initiative of the mobile industry to launch an ongoing security improvement program that is focused on mobile network infrastructure equipment. (GSMA n.d.). NESAS is part of the GSMA and together with 3GPP they have defined two elements for vendors:

1) security assessments of vendor development and product lifecycle processes
2) security evaluations of network products

The combination of both these activities defines and introduces a baseline security level that should be reached by the mobile industry. (GSMA n.d.)

The core itself is better secured now. By default, there is an assumption that protocols are encrypted by digital certificates. That ensures data privacy and makes spoofing remarkably harder, as communication parties can be identified now. However, encryption is not mandatory. Some troubleshooting reasons may require switching to unencrypted mode and that is fully working too if it is configured that network functions accept unencrypted and unsigned connections. An end user cannot see if MNO's core is running in unencrypted mode. This is important to notice because in 5G not only communication content or MNO's infrastructure is worth to protect, but there are visions that national power grid control would be carried out over sliced 5G and MECs. Compromised 5G core may also impact electricity delivery and that has consequences for the rest of society. While the 5G core has better built-in security mechanisms, the previous examples addressed that there are more boundary surfaces to protect, and their protection must be at the same level as the 5G core itself. (Int1, intv 2)

In the 5G core, network functions' connections protection by digital certificates requires certification authority (CA) to issue certificates. Because in this question connections are internal, the CA can be public or private. There is no one right answer should that be public or private. While public CA may provide some interoperability benefits in the future, a private CA provides better security, because CA infrastructure does not need to be exposed to the Internet at all.

Intv5 considered public CA very risky for an MNO, if the public root or sub CA gets compromised. It is much harder to compromise private root and sub CAs as they are not exposed to the Internet. The roaming case requires that MNOs need to trust each other's private PKI (root) certificates. It is a small extra work compared to use public PKI, but worth of doing.

Slice security is two-fold question. The slice itself, if it is properly configured without mistakes, should be secure. Home register permits a UE to use a certain slice and if just one slice is allowed, none of the interviewee saw any possibilities to break that. Isolation breakout scenarios requires that one device it allowed to use and attached to more than one slice. When multiple slices are allowed for one device, the 3GPP standard leaves the idea that certain traffic type belongs to certain slice. Slice parameters defines that, and UE should follow it. The problem is Android devices, because there are wide variation of OS releases and software, even malicious software can be freely installed. Other problematic devices are IoT devices, that may allow to upload custom firmware, again harmfully modified. All interviewees proposed almost the same solution to mitigate slice isolation breakout threat. In the first phase slice amount should be limited one slice per subscription, ie. SIM card. While one SIM and device is connected only one slice, there cannot be isolation breakout. That still brings biggest benefits of slicing, ie. guaranteed capacity, delays, and isolation. Once MNO's experience of slicing and possible threats increases, multiple slices per subscription can be considered. Intv5 also mentioned that PCF has capabilities to prevent certain isolation breakout threats, but it does not solve harmful IoT devices case. For them industrial control system would be needed. Deeper explanation of them was not covered during the interview, but Intv5 told where to find further reference and those mechanisms are covered by the researcher in Chapter 5.3.

Slicing breakout is not only something that happens at the UE side. At the core network side slicing customer needs to be isolated and limited in the relevant part of the network that are necessary to provide needed service. An important note from the Intv3 was that end-to-end slice separation and slicing breakout

prevention require in-depth solution before we can think to provide high security service and consumer Internet service over the sliced 5G network. From the commercial point of view, MNO and cost-efficiency point of view high security customer, both see this kind of resource efficiency very attractive, but Intv4 reminded that there must be enough time and resources for thorough testing prior new services are brought to production.

Network Exposure Function (NEF) is very interesting, because it is supposed to provide information exchange and service provisioning interface between 5G core and 3rd parties. It provides new possibilities, but also something very new to secure. For those reasons NEF is included in this thesis, but unfortunately NEF part was quite short in all interviews and not all interviewees were very familiar with it. Anyway, all have an assumption that NEF comes in production systems in the next 2-3 years, if the demand can be found. Intv1 reminded that there have been GSMA's one-API in 4G over 10 years, which basically can provide similar things already, but except of some pilots, no real need found. Intv2 saw NEF's potential opposite way and after SBA ramp up, it could be a solution for fast and flexible slicing provision needs. However, once NEF can allocate resources, it needs to be protected carefully by access control system and exposed only to actual business partners. NEF cannot be wide open to Internet.

Intv5 considered NEF to be good for security. It creates a single point to exchange information and also provides control who can access to SBA. The control is provided by role-based access control (RBAC), which defines users, roles, groups and information who can do what. A simple example how NEF with RBAC creates a slice:

1) call API to network element
2) establish a side session
3) connect NEF via API
4) NEF checks if user called an API is allowed to create a slice

It is not cybersecurity issue itself, but more like configuration issue, that NEF must not allow 3rd party to reserve more resources that are available. It must keep in mind as well, that there can be mobile virtual network operators (MVNO)

in MNO’s network, and they have contract-based resource allocations. That means that MVNO can use for example up to 10% of MNO’s network capacity, but if MVNO’s customers don’t need it, it is not statically reserved. MNO’s NEF must not over allocate resources and jeopardize MVNO’s capacity.

MEC security was another topic without good overview. All interviewees discussed it at a fairly general level because so many questions are still unclear. Most likely MECs are virtual hosts, but resources to be provided to customers are not clarified yet. They can be just computation capacity, OS on top of it or even applications. They all are possible, but security requirements and threats are totally different. It is self-evidence that MNO must prevent malicious code to be run on its MEC. However, it is not self-evident what actions should be accomplished to do that. Intv3 put it in words: “There is no such a service yet, so methods are not selected yet. Once we know what kind of MEC based service will be provided, there are plenty of security solutions available.”

Once dynamics increases in the SBA core, it will be too slow and complex task to maintain traditional firewalls and intruder detection systems (IDS). If NEF provisions a slice and allows access from and to customer network, dynamic firewall configuration needs to be solved. Or if a new MEC host created, normal traffic profile needs to be defined and updated as well. In a long term all interviewees think that some kind of artificial intelligent (AI) and machine learning (ML) based security solution is the only answer for those challenges. Again, the exact requirements of the security solution and type of new services to be provided in the first phase are unknown. AI and ML based security solutions are expensive, and MNO’s requirements needs to be carefully considered that right solution can be specified. Once requirements can be specified, it is time to evaluate choices, for example from what data and how well the solution can learn the baseline of the traffic. Now it seems to be too early with AI and ML based systems because requirements are still unclear. (Intv2).

7 DISCUSSION

All the interviewees had around 20 years' experience of the mobile technology but still questions and discussions of 5G's vulnerabilities and threats were difficult and remained quite general level. It is not a surprise, if compared what ENISA writes: "*The overarching nature of 5G, its complexity, the lack of information on existing deployments, the width and depth of existing specifications and the large number of potential stakeholders involved, makes the assessment of cyberthreats a difficult task."* (ENISA 2019, 10). This as a starting point created some challenge over the research questions, that were:

The primary question:

1. What does 5G network slicing mean for the operator's core network security?

Two secondary questions:

1. What kind of new threats will be arisen?
2. What actions should be considered to mitigate risks?

A comprehensive answer without any open questions left for all of these research questions would be a tremendous achievement. Nevertheless, results opened from the bottom to up means that risks need to be defined as well as actions to mitigate them. Generally, the risk management process consists of four things: identify, analyze, mitigate and monitor risks. Because of that, the risk management is information gathering and decision making. The focus is to understand feasible risks, classify and assess them and then determine their importance for business. (Kohnke et.al. 2016, c. 6.) ENISA has developed a methodology especially for 5G network risk assessment and that is presented in Figure 19.

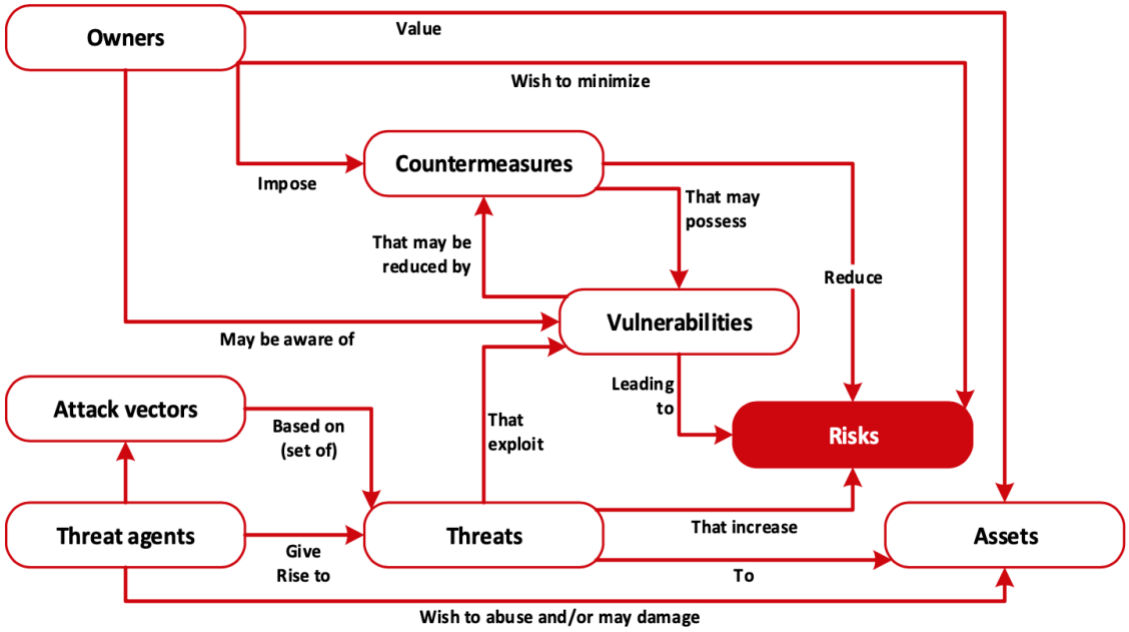


Figure 19. Risk assessment methodology, based on ISO 27005 (ENISA 2020, 12).

In this model threats have a central role in a risk assessment. By this methodology ENISA has identified assets, threats and threat agents. ENISA uses this methodology for the preparation of its annual Cyberthreat Landscape.

ENISA sees that “The 5G threat landscape may be useful to carry out detailed threat analyses and risk assessments for telecom operators and service providers according to their particular needs and mandate.” (ENISA 2019, 12). This methodology is applied while interview answers are decompiled to answers to research questions.

7.1 What kind of new threats will be arisen?

Threats are in central role in a risk assessment. ISO 27005 standard defines that risk emerge when: “Threats abuse vulnerabilities of assets to generate harm for the organisation” (ISO/IEC 27005:2018).

One concern are plans to use 5G networks to control other society critical services, like power grids and water systems. That creates indirect threat that is even bigger than threat against communications network. So far threats against communications networks have designated compromised user data, interrupted

communications or wrong billing, which are serious, but when compared to uncontrollable power outages or disrupted water systems, consequences are extended to everyone, even those who do not use 5G.

The high-level architecture in the 5G power grid controlling scenario involves MEC to accomplish control actions rapidly enough. This means operator core infrastructure to be located out of traditional physical security domain and physical core interfaces to be stretched at the customer premises. Interfaces are always causing threats, but physical location at the weak security area is definitely a new threat. From a network architectural point of view this means the user plane (UP) functions to be deployed closer to the edge of the operator network. As UP functions move to the edge, this pushes sensitive data to the network edge, therefore it increases the risk for sensitive data compromise. ENISA has created 5G threat taxonomy, which helps to categorize and find threats that 5G SBA, slicing and MEC raises as new threats. This thesis covers only new threats in the research scope field, but the reader must keep in mind that a full commercial mobile network is much more.

Physical attacks are possible mainly against MEC implementations that would be deployed at the customer premises. Other network functions of the SBA model are still in the operator's central data center and should be protected same way as in the past. However, MEC architecture means that core network interfaces are stretched at the customer premises and physical security must be ensured to mitigate this risk.

Legal threats can occur both slicing and network exposure function (NEF). In both cases threats are that contractual requirements and/or service level agreement (SLA) do not meet. In the slicing it is required to provide agreed service level (throughput and/or latency), but also limit overload and prevent resource exhausting. It is very similar to NEF, when it will be provided to 3rd parties. They must be able to reserve all the resources that have in the contract, but not more or it too may lead network resource exhausting.

Failures/malfunctions, **outages, disasters** are always possible with any kind of technology. While they are threats for the service continuity, they are not actual cybersecurity risks. Also, failures and outages are not such threats that come with new 5G service-based architecture, but they have always been present in the operator network.

Unintentional (accidental) **damages** are recognized threats by interviews. They may lead to cybersecurity threats if some vulnerability will expose because of human mistake. The mistake can be either inadequate design or misconfigured systems. Reasons behind both can be that personnel are not familiar enough with the new technology. At the same time there is time pressure to implement new systems in the production as soon as possible. That may lead insufficient testing, especially with the new technology. Information leakage belongs under unintentional damages as well. Especially with misconfigured NEF information leakage is possible, as NEF is intended to exchange information between operator core network functions and 3rd parties.

Eavesdropping / interception / hijacking are probably the most common threats expressed in the cybersecurity field. While they major role in the full 5G end-to-end security, including radio path, in the scope of this thesis, they form only minor threat. When considering 5G core and SBA, one must keep in mind that by design principles all the connections between network functions should be encrypted by digital certificates, and any kind of successful eavesdropping should be extremely difficult. Of course, if encryption is removed for example because of troubleshooting, this threat increases significantly. In this class and from the SBA point of view, abuse of roaming interconnections are potential threats, however, roaming was delimited out of this thesis.

Nefarious activity / abuse of assets covers such cybersecurity threats that are somehow valid for every IT-system threat analysis. In here it is appropriate to concentrate on those which especially are valid for interview results and 5G core. When compared with the interviews, essential threat classes from the NESAS report are abuse of information leakage, abuse of authentication, abuse of

virtualization mechanisms, denial of service, exploitation of software, and / or hardware vulnerabilities, malicious code or software and manipulation of hardware and software.

Abuse of information leakage has been present already prior 5G core and 5G SBA does not change the threat or its impact. However, it was raised in interviews that this must not be forgotten and part of the management and orchestration system must be monitoring and auditing system to detect possible breach and abuse. The threat is if that is overseen and not implemented properly.

Abuse of authentication can cause threat for NEF if authentication is poorly implemented and there is no additional access control to limit from where authentication requests can come. This class covers also abuse the credentials of existing accounts, but any modern authentication system should not use username – password -pair for authentication, but digital certificates.

Abuse of virtualization mechanisms is a threat against MEC systems if penetration to host management system has succeeded or NEF is improperly configured and allows abuse in that way. Depending on what is the gained access level, further threats can be network virtualization bypassing, virtualized host abuse, virtual machine manipulation and abuse of the cloud computational resources.

Denial of service threat is always present with IT-systems. Two functions covered in this thesis are relevant under this topic. Distributed denial of service (DDos) and flooding of core network components are valid threats against network exposure function (NEF) if that is without any protection open to Internet. Edge node overload is valid for MEC. In both cases potential impact is service unavailability.

Exploitation of software, and / or hardware vulnerabilities include application programming interface (API) threats. API exploitation against NEF or MEC may

lead to software tampering or even system execution hijack. Possible impacts are service unavailability and information integrity and destruction.

Malicious code or software with all its lower-level threats: injection attacks, viruses, malware, rootkits, rogue ware, worms/trojans and ransomware are valid for mobile edge computing (MEC) systems. It is inevitable that they need to be protected against all of these, but threat landscape depends on what kind of implementation and access is provided to MEC customer (=business partner) and what is the service to be provided to end users. Depending on how much the MEC customer has control over the MEC system, threats for the mobile operator's core are very different. Potential impacts vary also and can be service unavailability, information integrity or destruction and other software asset integrity or destruction.

Manipulation of hardware and software can mainly cause threat against MEC systems. This threat class includes false or rogue MEC gateway, which means that either rogue device should be able to be installed at the MEC location or manipulate valid device's configuration. Rogue device presumes also physical access to MEC location. In this class ENISA has mentioned manipulation of the network resource orchestrator and fake access network node as threats. While it is true that if that can happen, it is a threat. It is also very hard to accomplish and very unlikely, because there is very restricted access to SBA core. It came clear from interviews that even if many threat scenarios can be constructed, many of them are also simple to block, and this is one of them.

7.2 What actions should be considered to mitigate risks?

Threats are causing or increasing risks, as presented in Figure 19. When risks are identified, the next action is to mitigate them. A mobile network operator (MNO) has basically all these methods to mitigate risks: remove or lower the risk by improving systems, remove the system that causes the risk or accept the risk and its consequences if the risk realizes. In any case, risk assessment needs to be done to avoid uncontrolled risk scenarios. MNO must perform continues risk assessment and risk management including vetting and testing before on-

boarding 3rd party applications. MNO must also ensure that the security and privacy responsibilities are clearly defined.

Physical attacks risks against center core elements are mitigated as they used to be. The 5G SBA is not going to change anything. Centralized data centers should be well protected and physical access control to be based on electronic tags and personal pin codes. Those ensure that logs can be collected and afterwards can be seen, who entered to the data center and when. Mechanical keys should not be used, because they do not leave a log entry. A MEC server can be located at the customer premises. MEC can be part of the core network as it can run some core network functions to achieve low latency. That means it should be protected the same way as the central core. In practice it means that MEC systems should be placed in the closed cabinets and equipped with access control, meaning electronic tags and personal pin codes.

Legal threats are caused if technical protections against over subscriptions and/or SLAs cannot be met. There are two ways to mitigate this risk. Either technical systems to be able to ensure agreed service levels or possible SLA breach compensations to be kept at tolerable level. Legal risk should be evaluated as remaining risk after reasonable technical mitigation methods have been done.

Failures/malfunctions, **outages, disasters** mitigation requires validation, what happens if certain network function (NF) does not work as expected and how wide are the consequences. That is far simpler task than considering all the possible disasters what can threaten the service and cause a risk. Of course, for example fire and power outage types of disasters should be considered and recovery plans created for them, but malfunctions because of software bugs are harder to foresee. That leads to two approaches to mitigate this risk:

1) Minimizing the risk of software bugs and configurations incompatibility: carefully created testing plans, comprehensive tests with enough time, which also means time to fix found issues and complete re-testing to ensure that issue was really fixed

2) Creating a high availability (HA) environment: consider how critical is a certain network function (NF) for the service that is provided to customers and how much effort should be placed to make it redundant.

Both risk mitigation methods are highly based on business decisions. Some level testing is mandatory to be sure that there are no fatal or trivial errors, but endless testing cannot prevent to move on production. It must be accepted that testing is always somehow limited and live environment with thousands of users will reveal things that did not occur during the testing. This means that monitoring is an essential to mitigate this class risks and for the service continuity. Well defined monitoring can unveil threats / risks before they cause damage / breach.

Unintentional (accidental) **damages** are close to previous risks, but the difference is that in this class human staff is an active party of the risk. Configuration errors belong to this class and they should be taken seriously, as 5G is new even for experienced engineers. One of the best practices introduced by interviewees was not to trust only SBA network functions (NF) built-in security but protect them with external firewalls. While 3GPP design and NESAS's testing presume that properly configured NFs are secure without external security solutions, nothing protects them against configuration errors, unless there is additional security solution. Those additional security solutions are introduced in this thesis Chapter 5.

Eavesdropping / interception / hijacking risks are easily mitigated in the core network by acting as standard intends to do. It means that all connections between network functions (NF) are encrypted and signed by digital certificates. Digital certificates require public key infrastructure (PKI) to be used. In the PKI infrastructure there is certificate authority (CA) which issues digital certificates to devices, or to NFs in this case. CA guarantees the validity of the certificates and therefore network functions identities as well. (SSH Academy N.d.). CA's functionality and security is essential for the 5G core network. Because there is no strict reason to use public CA, maximum security can be achieved to use internal CA that cannot be accessed from Internet. This restriction prevents outside DoS type attacks against the PKI.

Nefarious activity / abuse of assets threats are similar to any IT-systems and thus mitigation methods too. In the past there have been specific telco protocols, but in the SBA, everything is on top of internet protocol (IP). This means that 5G core needs to be separated to different security zones same way as general data centers. Firewall placement design is like data centers and the principle how security zones separation should be performed is presented in Figure 16. Each zone should be separated by border firewalls which create multi-layer isolation, as described in Chapter 5.5. That kind of separation helps to prevent information leakage as there are separate devices and rules limiting allowed connections.

Abuse of authentication can be mitigated in NEF case by placing firewalls in front of authentication server and restricting access only to known parties' IP addresses. In all authentication cases firewall should do rate limiting to prevent brute force attack. Rate limiting also reduces effectivity of the basic denial-of-service (DoS) attacks. MEC authentication contains at least two dimensions. The first one is to authenticate applications running on MEC platform. There should be in place a system that ensures that all running applications have been digitally authorized to be executed. Abuse of such an authentication system should not be easy, because it should be available only in an operator's MEC management network. Another dimension is to authenticate MEC application users or clients if they are machines. Because MECs work in the mobile world, there is Mobile Station International Subscriber Directory Number (MSISDN), ie. mobile phone number, available for authentication. Every mobile subscriber must have that to be able to join the network and it also must be valid from the MNO. Human users may can use username/password-pair for authentication, but for machines that is not very practical. Because MSIDSN is available in every mobile subscriber, there are very few reasons, why it would not be used to client authentication to MEC resources, at least abuse is very hard when MSISDN is required.

Abuse of virtualization mechanisms has link to authentication. Every application running in MEC should be digitally signed and permitted to be executed. Unsigned or un-permitted application needs to be prevented from running by the

virtualization system. That prevents abuse of cloud computation resources. Virtual machine manipulation and host abuse can be mitigated by removing local administration accounts and allowing only personal centralized accounts that usage will also be logged. If automated management systems are used for the virtualization platform operation and maintenance, no static username/password-pairs are to be used, but digital certificates and encrypted connections instead.

Exploitation of software, and / or hardware vulnerabilities is already somehow mitigated when connections to MEC are limited to allowed parties by MSISDN. Anyhow, in this case too, allowed party may still accomplish hostile exploitation. In Chapter 5 described machine learning (ML) based security solutions would be effective to mitigate this risk. ML-based system to be taught of the normal traffic patterns and exceptions from that will raise an alarm to SIEM or situational awareness system.

Malicious code or software are threats against virtualized hosts on MEC platforms. Threats in this category are very common in any server system, like injection attacks, viruses, malware and trojans. Mitigation methods are also similar to computer cybersecurity field, keep systems updated, use anti-virus software, lock down unnecessary services, run software with minimal privileges and run penetration testing / security scanning tool for time to time.

Summary of all risk mitigation from ENISA report is in line with the interviewees' views. It is important to understand what is going to be implemented and enough time for testing. A systematic test plan must occur for new features and separate security test plan to run after each patch or update. A risk reporting system is also crucial, that anyone noticing a risk can report it. Once the risk has been reported, there must be a person who is responsible to do further analysis for the risk and advance the risk to the correct asset owner. It is important that the risk reported must not be directly in charge of further actions of the risk, because it might cause a threshold to report identified risk.

7.3 What does 5G network slicing mean for the operator's core network security?

Slicing means that physical network resource divided into two or more logical resources. It is like what happened to servers when they moved to virtualized on top of VMware or other hypervisors, or when one Ethernet switch began to serve multiple separate LAN segments with virtual-LAN (vlan) technology. Those technologies did not weaken security but changed it and from some point of views the security also increased. Virtualized server is quicker to patch and if one installation became compromised, it is easy to revert to non-compromised snapshot or fresh installation. In the Ethernet switches vlans makes it cost efficient to deploy separate security zones from one switch rather than considering if everything could put in the same LAN segment. 5G core slicing has many same aspects. It is new way to share resources, new techniques to learn, new aspects to keep in mind, but none of them are less secure than the previous mobile core. Of course, poor implementation can cause security breaches, just like an inadequate vlan configuration in the Ethernet switch can cause.

Especially enterprise customers are looking for dedicated and guaranteed transfer rates. They might have production control or security monitoring systems that must not suffer lack of transfer capacity at any time. The answer for those needs is most of times radio access network (RAN) slicing. RAN slice means that certain amount of base station bandwidth will be dedicated to one customer. That guarantees that the critical traffic does not suffer occasional radio congestions. The core network slicing is not needed at all if the requirement is simple as that. The core slicing comes in the picture, if the radio path guarantee is not enough, but end-to-end delays need also to be minimized.

Both 3GPP recommendations and all interviewees said that core slicing should start from basics. 3GPP has defined three main classes (eMBB, mMTC and uRLLC). NESAS and other organizations have studied 3GPP's basic models and threat landscape is discovered. Starting from basics means that there are plenty of guidance available from vendors and they can also carry part of the risk. In a long run three static basic slices would not be enough. The risk would be bigger

when basic and generic slice types are not enough. Dynamic slice provisioning will be needed because of business demand. It must be ensured that over provisioning is prohibited as well as any kind of un-authorized provisioning. When some slice is used for the society critical service, slicing breakout also must not happen. It is more than obvious that all reasonable actions have been taken to prevent said disasters, but still some unknown vulnerabilities are left. Understanding that and remembering how complex system the 5G network is, machine learning (ML) based security solution is expected to be great improvement. As interviewee2 reported in line with Suomalainen et.al.'s report, the question is about balance of the costs and ML security solution effectiveness. It is right time for the ML based security solution, once its costs are lower than risks that it is mitigating. Nevertheless, some kind of the situational awareness system is mandatory to visualize how resources are utilized and alarms to be raised if abnormal resource usage occurs.

NEF security is a twofold case to solve. It provides the provision service for a 3rd party as described in the previous chapter, but it can also provide information from the 5G network to 3rd parties. It should be self-evident, that there needs to be policies and rules, what information to provide, but also controls that they are applied, and no data leakage happen. Once NEF will be implemented, also with that should be started with basics. If NEF will be used to offer some information form the 5G network, then for example simple location information. If NEF will be used as provisioning interface, maybe it would be good to allow only fixed characteristics slices in the beginning. But whatever the service will be, abnormalities and even suspicious attempts are important to be notified immediately. If they remain unseen for a too long time, the risk for major data leakage occurs.

7.4 Limitation and reflection

Because the research was conducted before the technical prerequisites were ready, no other methodology than expert interviews would provide the better real-life results. Perhaps a bit more independent researcher would obtain expert interviewees from all major mobile operators and that would have been provide

somehow better perspective over the industry. Interview questions, or even just topics to discuss like in this thesis, are always subjective. Some other researcher might emphasize experts' comments slightly other way. Even though in this thesis the researcher reflected comments against the literature review to make sure that answers are somehow in line with the common understanding.

The SBA core is covered by many reliable research and other publications, and all interviewees were familiar with the topic. MEC was much more at the theoretical level. Especially lack of real-life applications made it difficult to cover real-life cybersecurity threats. Now it was much speculation, if the deployment is like this, then risks are these. In this kind of master's thesis, there would easily come a temptation to give recommendations for a good security practice regardless of the work loads. However, one must keep in mind that MEC as well as any IT system is to be set up to produce some service to customer. So once the service is known, the security can be designed in detail.

The network expose function (NEF) is interesting function needed for large scale and automated network slicing. It is something that is needed to fulfill all scenarios of general-purpose network, as 5G aimed to be. Without NEF slicing is not as flexible and fast to provision as it need to be. NEF was covered here based on the literature, but real-life implementations and security point of views remained to the future. Protecting NEF and its services might be worth of further master thesis studies after a year or two.

The researcher has been in the telecommunication industry over 25 years. Some explanation of the supporting technologies, like PKI infrastructure, felt little bit back to basics. It is so fundamental part of security solutions that it should be known even without explanation in this kind of report. However, that is not a case with master's thesis, but PKI needs to be explained. Because of that, this kind of basics were first written and supporting source was searched afterwards. This is not a best method, because most likely some aspect is lost. However, it saves a lot of time to use essential topics.

One contradiction can be expressed of the purpose of the thesis and given grade objectives. The ultimate target is to produce new information for the whole industry. In the cybersecurity area gaining such an expertise requires years and years working experience and specialization to certain area. It would be quite an achievement from a student of couple tens course credits of the cybersecurity area to present something new to whole industry. That even though each student should have some professional expertise before master's thesis project. After all, a thesis is a written report, which should demonstrate that the student can understand and process wide information entities and correlations between them. Master's thesis is the end of studies, but the beginning of the learning, if a new profession was the target before studies.

8 CONCLUSION

The outcome is that core network slicing itself does not collapse the security of the 5G core. Starting from the basics and then well phased growth also makes it easier to keep core secured. It is not useful to add services like MEC or NEF if there is no business demand, in that way they only add threat landscape. It is better to find use cases, why they would be needed and then implement them to fulfill that particular demand. It is easier to do risk- and threat analysis when the use case is known. That approach allows to crop all other use cases, but actual needed ones and shut down unnecessary services or restrict allowed connections, ie. do the traditional system hardening. All that also means a landscape with smaller threat.

8.1 Summary of the study

The journey from an idea to completed thesis took roughly a year. In the late 2020 the 5G core slicing was a roadmap item of the technology vendors. All parties were preparing to the full service-based architecture (SBA) core, attach 5G base station to it and soon start 5G radio network slicing. In that time core network slicing and additional services close to it, like mobile edge computing (MEC) and network expose function (NEF) felt logical next step and worth of deep studying.

On autumn 2021 SBA core and radio network slicing is just about to get their commercial launch. It also means that core slicing, MEC and NEF are not available yet, but they are next in the list. Because of that, this study relied on expert interviews to find out, what kind of secure sliceable SBA core will be, what is MEC for and how the NEF should be implemented. Those services are not ready in the operator network, and to keep this thesis more than just a literature review, expert interviews provided a good practical approach for the research.

8.2 Recommendations

Based on the interviews, topics were more than relevant for an MNO and discussions were interesting and encouraged to new thinking at both sides. When discussions proceeded deeper from the surface, many times answers were something like: “interesting, that should be tested / learned more when that is available”. It means that there are good starting points to discover how current situation awareness systems would need be further developed to be effective with dynamic slicing and NEF. Same applies to MEC, once known what kind of services are to be provided, effective real-life security can be designed.

8.3 Contribution to research and practice

The main purpose of this master’s thesis was to increase the researcher’s own knowledge of the 5G core slicing and new services that can offer and of course, how to do them secure. The researcher and team he works in, need the thesis report knowledge when new enterprise customer services are developed. The research construed what is essential in the area, what is the readiness of the technology and what are their correlations. For example, NEF is not needed if there are only few static slices. Another example would be that minimized delays would require MEC to be placed at the customer site, but then MEC needs physical protection. That leads to the question, would that be worth of the last couple milliseconds. Most likely autonomous driving requires it, but it might be difficult to find any other services before autonomous driving is ready for general use. This master’s thesis gives perspective to consider choices and variables.

Costs was not in the scope at all, but in the workday life the researcher must consider technology costs all the time and that is essential for his team and commissioner too.

Even though threats and security solutions discovered here are not something very special or new to experienced mobile technology experts, findings can be valuable to business side people. They might not know very well all technology details of the 5G core slicing security and from this report they can be found in one place.

REFERENCES

3GPP 5G Security. 2018. 3GPP. WWW document. Available at: https://www.3gpp.org/news-events/1975-sec_5g [Accessed 3 May 2021].

3GPP FAQs. N.d. 3GPP. WWW document. Available at: https://www.3gpp.org/about-3gpp/3gpp-faqs#Specification [Accessed 4 August 2021].

3GPP Releases. N.d. 3GPP. WWW document. Available at: https://www.3gpp.org/specifications/67-releases [Accessed 13 January 2021].

3GPP Release 15. 2019. 3GPP. WWW document. Available at: https://www.3gpp.org/release-15 [Accessed 14 January 2021].

3GPP Release 17. 2020. 3GPP. WWW document. Available at: https://www.3gpp.org/release-17 [Accessed 4 August 2021].

3GPP Release 17 timeline agreed. 2020. 3GPP. WWW document. Available at: https://www.3gpp.org/news-events/2145-rel-17_newtimeline [Accessed 4 August 2021].

3GPP Specifications Groups. 2021. WWW document. Available at: https://www.3gpp.org/specifications-groups [Accessed 4 August 2021].

5G Americas. 2019. The Evolution of Security in 5G, A “Slice” of Mobile Threats. White Paper. Available at: https://www.5gamericas.org/wp-content/uploads/2019/08/5G-Security-White-Paper-07-26-19-FINAL.pdf [Accessed 5 September 2021].

5G Americas. 2021. 3GPP releases 16 & 17 & Beyond. White paper. Available at: https://www.5gamericas.org/wp-content/uploads/2021/01/InDesign-3GPP-Rel-16-17-2021.pdf [Accessed 4 August 2021].

AdaptiveMobile Security. 2021. A Slice in Time: Slicing Security in 5G Core Networks – WHITE PAPER. WWW document. Available at: https://f.hubspotusercontent10.net/hubfs/8487362/AMS_Slicing_Security_in_5G_Core_Networks_Whitepaper_v1.00.pdf?utm_campaign=White%20paper%20-%20Telco%20Security&utm_medium=email&_hsmi=117151103&_hsenc=p2ANqtz-8wCsieBQBLX1eqdLz4i9MEJ-Rew-Mzzd1-TLbLMzudKF91vwbOLXrmGVwIsjd4uuoigdTjWFnAVEtMp88q4mhrxHZ5wQ&utm_content=117151103&utm_source=hs_automation [Accessed 15 April 2021].

Ahmad, I., Kumar, T., Liyanage, M., Okwuibe, J., Ylianttila, M. & Gurtov, A. Overview of 5G security challenges and solutions. IEEE communications standards magazine. 2019. Available at: http://jultika.oulu.fi/Record/nbnfi-fe201902124647 [Accessed 26 March 2021].

Alastalo, M., Åkerman, M. & Vaittinen, T. 2017. Asiantuntijahaastattelu. In Hyvärinen, M., Nikander, P., & Ruusuvuori, J. (eds.) Tutkimushaastattelun käsikirja. Tampere: Vastapaino, 181-197. Available at: https://kaakkuri.finna.fi/ [Accessed 19 July 2021].

Baškarada, S. Qualitative Case Study Guidelines. 2013. WWW document. Available at: https://apps.dtic.mil/sti/pdfs/ADA594462.pdf [Accessed 7 March 2021].

Council of the European Union. 2019. Law enforcement and judicial aspects related to 5G. WWW document. Available at: https://data.consilium.europa.eu/doc/document/ST-8983-2019-INIT/en/pdf [Accessed 24 August 2021].

ENISA. 2019. Threat Landscape for 5G Networks Report. WWW document. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-for-5g-networks [Accessed 5 August 2021].

ENISA. 2020 Threat Landscape for 5G Networks Report. WWW document. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-for-5g-networks [Accessed 5 August 2021].

Golic, J., Zuo, M., Wang, K., Zhuang, X., Qi, M., Hartmann, C., Kneppers, M., Yogendra, S. & Rosalia, D'A. 2018. Security Aspects of Network Capabilities Exposure in 5G. NGMN Alliance. WWW document. Available at: https://www.ngmn.org/wp-content/uploads/Publications/2018/180921_NGMN-NCEsec_white_paper_v1.0.pdf [Accessed 17 Marc 2021].

Gonzales, A. J., Ordonez-Lucena, J., Helvik, B. E., Nencioni, G., Xie, M., Lopez, D. R. & Gronsund, P. N.d. The Isolation Concept in the 5G Network Slicing. WWW document. Available at: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2685620/Postcopy_2020-04-30_1570629806-2.pdf?sequence=1 [Accessed 2 March 2021].

GSMA. 2021. Securing the 5G Era. WWW document. Available at: https://www.gsma.com/security/securing-the-5g-era/ [Accessed 3 May 2021].

GSMA. N.d. NESAS FAQs. WWW document. Available at: https://www.gsma.com/security/nesas-faqs/ [Accessed 21 July 2021].

GSM Association. 2021. E2E Network Slicing Architecture, version 1.0. Official Document NG.127, Available at: https://www.gsma.com/newsroom/wp-content/uploads//NG.127-v1.0-1.pdf [Accessed 22 June 2021].

Guan, W., Wang, L. & Lu, Z. 2018. A Service-Oriented Deployment Policy of End-to-End Network Slicing Based on Complex Network Theory. WWW document. Available at: https://www.researchgate.net/publication/324175599_A_Service-Oriented_Deployment_Policy_of_End-to-

End_Network_Slicing_Based_on_Complex_Network_Theory [Accessed 14 August 2021].

Harel, R., & Babbage, S. 2016. 5G security - Package 3: Mobile Edge Computing. NGMN Alliance. WWW document. Available at: https://www.ngmn.org/wp-content/uploads/Publications/2016/161028_NGMN-5G_Security_MEC_ConsistentUExp_v1.3_final.pdf [Accessed 8 April 2021].

Harel, R. & Babbage, S. 2016. 5G security recommendations Package #2: Network Slicing. NGMN Alliance. WWW document. Available at: https://www.ngmn.org/wp-content/uploads/Publications/2016/160429_NGMN_5G_Security_Network_Slicing_v1_0.pdf [Accessed 27 March 2021].

Hirsjärvi, S. & Hurme, H. 2015. Tutkimushaastattelu: Teemahaastattelun teoria ja käytäntö. Helsinki: Gaudeamus.

Hyvärinen, M., Nikander, P. & Ruusuvuori, J. 2017. Tutkimushaastattelun käsikirja. Tampere: Vastapaino. Available at: https://kaakkuri.finna.fi/ [Accessed 3 July 2021].

Ikonen, H-M. 2017. Puhelinhaastattelu. In Hyvärinen, M., Nikander, P., & Ruusuvuori, J. (eds.) Tutkimushaastattelun käsikirja. Tampere: Vastapaino, 230-243. Available at: https://kaakkuri.finna.fi/ [Accessed 4 July 2021].

ISO/IEC 27005:2018. Information technology — Security techniques — Information security risk management. Non-public.

Ivezis, M. 2020 Introduction to 5G Core Service-Based Architecture (SBA) Components. WWW document. Available at: https://5g.security/5g-technology/5g-core-sba-components-architecture/ [Accessed 5 March 2021].

Kekki, S., Featherstone, W., Fang, Y., Kuure, P., Li, A., Ranjan, A., Purkayastha, D., Jiangping, F., Frydman, D., Verimn, G., Wen, K., Kim, K., Odgers, A., Conteras, L. & Scarpina, S. 2018. MEC in 5G networks. ETSI White Paper No. 28. Available at: https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp28_mec_in_5G_FINAL.pdf [Accessed 22 March 2021].

Kler, Y. 2020. 5G Security: Mobile Edge & Network Slicing. Huawei. Non-public.

Kohn, L. & Christiaens, W. 2018. KCE Process Book. Available at: https://processbook.kce.fgov.be/node/1 [Accessed 29 July 2021].

Kohnke A., Shoemaker D. & Sigler K. 2016. The Complete Guide to Cybersecurity Risks and Controls. Boca Raton, FL: CRC Press.

Kokkonen, J., 2020. Elisa testaa Pohjoismaiden ensimmäistä 5G-standalone-verkkoa yhdessä Ericssonin kanssa. *iO.TECH*. WWW Document. Available at:

https://www.io-tech.fi/uutinen/elisa-testaa-pohjoismaiden-ensimmaista-5g-standalone-verkkoa-yhdessa-ericssonin-kanssa/ [Accessed 4 September 2021].

Kurtz F., Bektas C., Dorsch N. & Wietfeld C. 2018. Network Slicing for Critical Communications in Shared 5G Infrastructures - An Empirical Evaluation: https://www.kn.e-technik.tu-dortmund.de/.cni-bibliography/cnidoc/Kurtz2018network.pdf [Accessed 2 March 2021].

Lampi, M. 2020. Web-based APIs in digital platform innovation. Master's thesis, University of Jyväskylä. Available at: http://urn.fi/URN:NBN:fi:jyu-202012157105 [Accessed 24 July 2021].

Lehto, T. 2018. Elisa avasi ensimmäiset 5g-verkot – 2,2 gigabittiä sekunnissa – tosin päätelaitteet puuttuvat vielä. *Tekniikka & Talous*, Available at: https://www.tekniikkatalous.fi/uutiset/elisa-avasi-ensimmaiset-5g-verkot-2-2-gigabittia-sekunnissa-tosin-paatelaitteet-puuttuvat-viela/18ffea6d-084e-3895-b009-718dbd6e48ad [Accessed 27 January 2021].

Muikku, J-M. 2020. Improving Cyber Security Situational Awareness with Log and Network Security Monitoring. Master's thesis, South-Eastern Finland University of Applied Sciences. Available at: http://urn.fi/URN:NBN:fi:amk-2020092220624 [Accessed 28 August 2021].

NGMN Alliance. 2015. 5G White Paper. WWW document. Available at: https://www.ngmn.org/wp-content/uploads/NGMN_5G_White_Paper_V1_0.pdf [Accessed 14 Marc 2021].

Penttinen, J. T. J. 2019. 5G Explained: Security and Deployment of Advanced Mobile Communications. Atlanta, Georgia, USA: John Wiley & Sons, Incorporated.

Positive Technologies. 2020. 5G Standalone core security research. WWW document. Available at: https://positive-tech.com/storage/articles/5g-sa-core-security-research/5g-sa-core-security-research.pdf [Accessed 15 August 2021].

Putkonen, J. 2019. 5G for Future Industrial Internet. 5GMomentum verkostoitumistilaisuus 6.11.2019. Available at: https://www.traficom.fi/sites/default/files/media/file/Putkonen%20Jyri_5GMomentum_5GIIoT_JPutkonen_20191029.pdf [Accessed 4 August 2021].

Rommer, S., Hedman, P., Olsson, M., Frid, L., Sultana, S. & Mulligan, C. 2019. 5G Core Networks: Powering Digitalization. Elsevier Science & Technlogy. 3GPP release 15 and 16.

Ruusuvuori J. & Nikander P. 2017. Haastatteluaineiston litterointi. In Tutkimushaastattelun käsikirja. Tampere: Vastapaino. Available at: https://kaakkuri.finna.fi/ [Accessed 8 July 2021].

Spencer L., Ritchie J., Ormston R. & O'Connor W. 2014. Qualitative Research Practice. London: Natcen: SAGE Publications.

SSH Academy. N.d. What is PKI (Public Key Infrastructure)? WWW document. Available at: https://www.ssh.com/academy/pki [Accessed 29 August 2021].

Stake, R. E.1995. The art of case study research. Thousand Oaks, CA: SAGE Publications.

Sun, T. & Wang, D. 2019. Service-Based Architecture in 5G: Case Study and Deployment Recommendations. NGMN Alliance. WWW document. Available at: https://www.ngmn.org/wp-content/uploads/Publications/2019/190919-NGMN_Service-BasedArchitecturein5GCaseStudyandDeploymentRecommendations-v2.4.pdf [Accessed 15 March 2021].

Suomalainen, J., Juhola, A., Shahabuddin, S., Mämmelä, A. & Ahmad, I. 2020. Machine Learning Threatens 5G Security. IEEE Access, 8, 190822 - 190842. WWW document. Available at: https://doi.org/10.1109/ACCESS.2020.3031966 [Accessed 30 August 2021].

Ta-Hao T., Tsung-Nan L., Shan-Hsiang S. & Yu-Wei C. 2019. Guidelines for 5G End to End Architecture and Security Issues. WWW document. Available at: https://arxiv.org/pdf/1912.10318.pdf [Accessed 2 Feb 2021].

Tovinger, T, Cornily, J-M., Gardella, M., Shan, C., Ai, C., Andrianov, A., Chou, J., Groenendijk, J., Kai, Z., Lan, Z., Sun, X., Wan, W., Weihong, Z. & Yao, Y. 2018. Management, Orchestration and Charging in the New Era. WWW document. Available at: https://www.riverpublishers.com/journal/journal_articles/RP_Journal_2245-800X_6110.pdf [Accessed 13 August 2021].

Traficom. 2021. 5G projekteja ja kokeiluja Suomessa. WWW document. Available at: https://www.traficom.fi/fi/viestinta/viestintaverkot/5g-projekteja-ja-kokeiluja-suomessa [Accessed 4 September 2021].

Yazan, B. 2015. Three Approaches to Case Study Methods in Education: Yin, Merriam, and Stake. The Qualitative Report, 20(2). WWW document. Available at: https://nsuworks.nova.edu/tqr/vol20/iss2/12 [Accessed 13 March 2021].

Yin, R. K. 2002. Case study research: Design and methods. Thousand Oaks, CA: SAGE Publications.

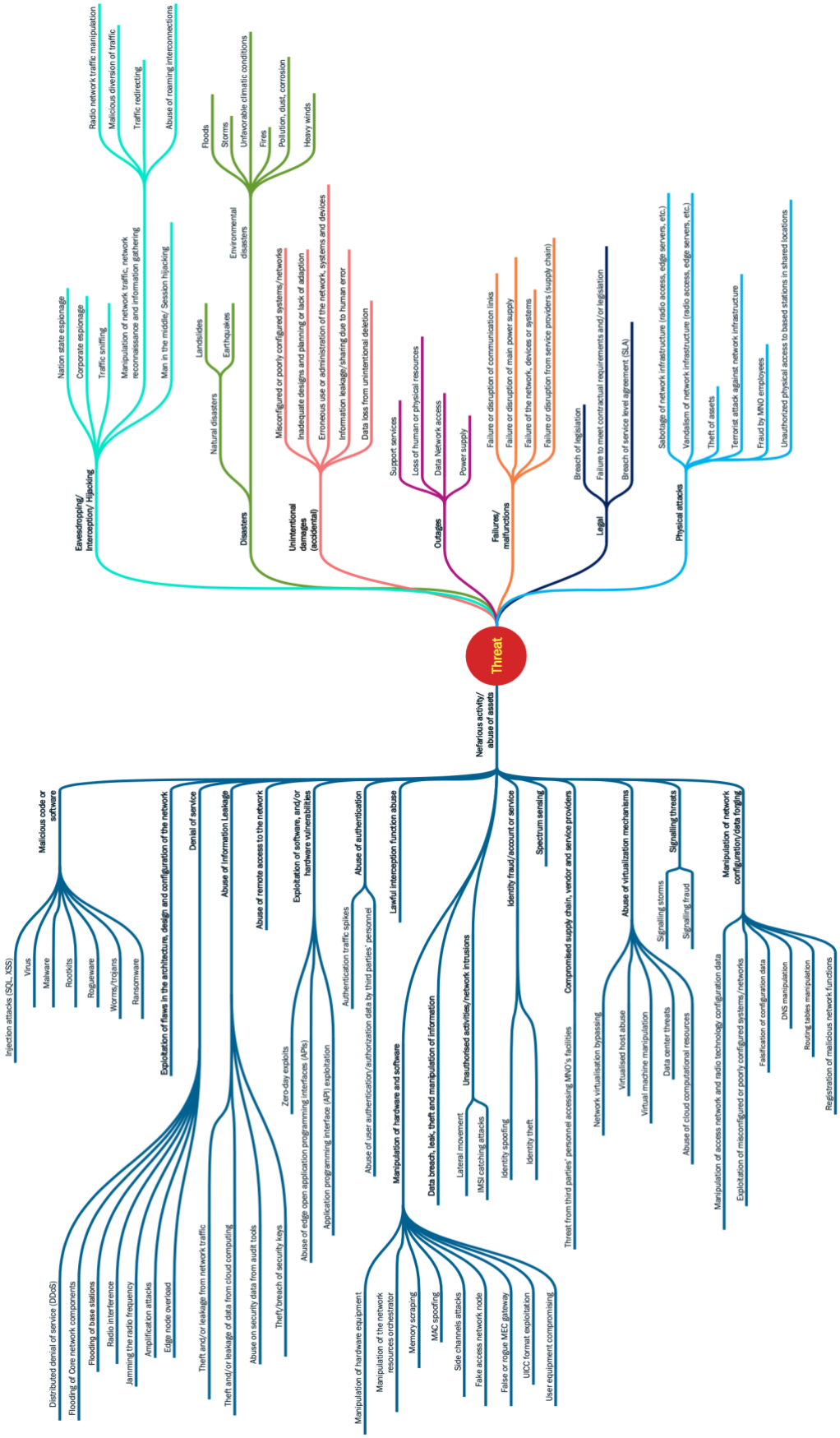
Appendix 1/3

LIST OF FIGURES

ACRONYMS

3GPP	3rd Generation Partnership Project
ADMF	Administration Function
AF	Application Function
AEF	Application Enablement Function
AMF	Access (and Mobility) Management Function
API	Application Programming Interface
ASN	Automated Sliced Network
AUSF	Authentication Server Function
CA	Certificate Authority
CHOP	Configuration, Healing, Optimisation, and Protection
CPE	Customer Premise Equipment
CPES	Control Plane Exposure Service
CPS	Control Plane Service
CSP	Communication Service Provider
CU	Centralized Unit
CUPS	Control/User Plane Separation
DN	Data Network
DNAI	Data Network Access Identifier
DNN	Data Network Name
DoS	Denial-of-Service
DU	Distributed Unit
E2E	End-to-end
eMBB	enhanced Mobile Broadband
eSCP	enhanced Service Communication Proxy
ETSI	European Telecommunications Standards Institute
GPSI	Generic Public Subscription Identifier
GST	Generic Slice Template
hSEPP	home Security Edge Protection Proxy
IoT	Internet of Things
LADN	Local Area Data Network
LMF	Location Management Function
MANO	Management and Orchestration
MEC	Multi-access Edge Computing (in ETSI nomenclature) / Mobile Edge Computing (in NGMN nomenclature)
mMTC	massive Machine-Type Communications
MNO	Mobile Network Operator
MSISDN	Mobile Station International Subscriber Directory Number
MSP	Mobile Service Provider
MVNO	Mobile Virtual Network Operator
NAS	Non-Access Stratum
NDL	Network Data Layer
NEF	Network Exposure Function
NEST	Network Slice Type
NF	Network Function
NFV	Network Function Virtualization
NFVI	Network Function Virtualization Infrastructure

NSI	Network Slice Instance
NSP	Network Slice (/Service) Provider
NSSAA	Network Slice Specific Authentication & Authorization
NSSAAF	Network Slice Specific Authentication & Authorization Function
NSSAI	Network Slice Selection Assistance Information
NSSF	Network Slice Selection Function
NSM	Network Management System
NSSMF	Network Slice Subnet Management Function
NR	New Radio
NRF	Network Repository Function
PCF	Policy Control Function
PGW	Packet gateway
PGW-UP	Packet gateway user plane
PDN	Packet Data Network
PDU	Protocol Data Unit
PKI	Public Key Infrastructure
RAN	Radio Access Network
RAT	Radio Access Technology
RBAC	Role-Based Access Control
SBA	Service Based Architecture
SCP	Service Communication Proxy
SCEF	Service Capability Exposure Function
SD	Slice Differentiator
SDN	Software Defined Network(ing)
SDO	Standards Development Organisation
SEPP	Security Edge Protection Proxy
SGW	Serving gateway
SGW-UP	Serving gateway user plane
SLA	Service Level Agreement
SMF	Session Management Function
SMP	Service Messaging Platform
S-NSSAI	Single Network Slice Selection Assistance Information / Subscribed Network Slice Selection Assistance Information
SP	Service Provider
SSC	Session and Service Continuity
SST	Slice Service Type
TR	Technical Report
TS	Technical Specification
TSG	Technical Specification Group
UE	User Equipment
UDM	Unified Data Management
UDR	Unified Data Repository
URLLC	Ultra-Reliable Low-Latency Communication
UP	User Plane
UPF	User Plane Function
UPS	User Plane Service
VNF	Virtual Network Function



(ENISA 2020, 128)