5G security recommendations
Package #2: Network Slicing

by NGMN Alliance

Version:	**1.0**
Date:	**27-April-2016**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**
Authorised Recipients: (for CR documents only)	

Project:	**NGMN - 5G Security**
Editor / Submitter:	**Rémy HAREL, Steve BABBAGE**
Contributors:	**NGMN 5G security group**
Approved by / Date:	**NGMN Board, 29th April 2016**

Commercial Address:
ngmn Ltd.,
Großer Hasenpfad 30 • 60598 Frankfurt • Germany

Phone +49 69/9 07 49 98-04 • Fax +49 69/9 07 49 98-41

Registered Office:
ngmn Ltd.,
Reading Bridge House • George Street • Reading • Berkshire RG1 8LS • UK
Company registered in England and Wales n. 5932387, VAT Number: GB 918713901

Abstract: Short introduction and purpose of document

The purpose of the NGMN 5G security group, which is a sub-group of the NGMN P1 5G Architecture group, is to identify new threats and security issues that may arise with 5G. The work of the group aims at informing the 5G community, and especially SDOs (like 3GPP) of potential problems in 5G, that we should pay attention to.
This working group does not make requirements, just recommendations that people should have in mind when designing 5G networks.

The group will release several documents with different focus; its first package focused on Improving the Access Network and also identified DoS attack scenarios in a 5G context. This second document focuses on security threats or flaws that could emerge through the Network Slicing use in 5G. While network slicing might not be the only approach that will be used in implementing 5G networks, it is envisioned that network slicing will have a very important role. Network Slicing is generally tied to Virtualization; some security concepts described here might be linked, applicable or even dependant on Virtualization.

It also must be noted that the Network slicing concept has many possible approaches and applications. The main architectural reference for this security study was the NGMN E2E architecture group's document named "Description of Network Slicing Concept" [1]

Document History

Date	Version	Author	Changes
Apr 14, 2016	V 0.1	Remy HAREL (ORANGE)	Creation of the document, import of validated key issues
April 19, 2016	V0.2	All	minors corrections
April 19, 2016	V0.3	Remy HAREL (ORANGE)	Addition of a note in key issue 10 for ME/UE clarification
April 21, 2016	V0.4	Remy HAREL (ORANGE)	Editorial corrections
April 27, 2016	V0.5	Remy HAREL (ORANGE)	Editorial corrections
April 29, 2016	V1.0	Klaus Moschner (NGMN Office)	Final Version

REFERENCES

Ref	Document
1	NGMN “Description of Network Slicing Concept” v1.0
2	NGMN “5G Security Recommendations – Package #1” (to be published)

GLOSSARY

Abbreviation	Meaning
3GPP	3rd Generation Partnership Project
AuC	Authentication Server
CN	Core Network
DNS	Domain Name Service
DoS	Denial of Service
DDoS	Distributed Denial of Service
E2E	End to End
HSS	Home Subscriber Server
IP	Internet Protocol
ME	Mobile Equipment (from 3GPP nomenclature)
MEC	Mobile Edge Computing
QoS	Quality of Service
RAN	Radio Access Network
SDN	Software Defined Network
UE	User Equipment (from 3GPP nomenclature)
VNF	Virtualized Network Function
VPN	Virtual Private Network

Contents

1 INTRODUCTION

1.1 Assumptions and definitions

We assume that “network slices” do not extend into mobile devices. This does not, however, preclude the possibility that one UE is attached to multiple slices simultaneously.

The concept of isolation between network slices is significant, and will be referred to in the sections that follow. It is important to understand, though, that there are several possible senses of “isolation between network slices”, e.g.:

- Ring-fencing resources for slice A, so that no matter how busy slice B becomes, it cannot expand to exhaust slice A’s resources.
- Not supporting communication between slices.
- Protecting against someone who is deliberately trying to “hack through the walls” between slices.

2 KEY ISSUES

2.1 Key Issue 1: Controlling Inter-Network Slices Communications

2.1.1 Description

While not a concrete entity but rather a logical grouping of inter-working components (functions, etc.), a network slice will have an overall ingress/egress communication on the user-plane.

We similarly anticipate that functions, used in network slices, will have similar user-plane communications characteristics.

In addition, both network slices and their components will have signaling and management plane communications requirements (with each other, and with elements such as an orchestrator).

As these communications are important to the functioning of the slice and its role in providing specific services to the operator network, we should protect against their disruption and ensure that there are no undesired communications (which might have the potential to disrupt the functioning of the slice).

2.1.2 Recommendation

- All communications between slices and functions should have a mechanism to control their transmission and receipt, in order to ensure operation within expected parameters and security.
- For all interfaces between slices and functions, consider defining security mechanisms to ensure operations within expected parameters and security needs of the operator.

2.2 Key Issue 2: instantiation time Impersonation attacks against Network Slice Manager or Host (physical) platforms within an operator network

2.2.1 Description

Assume that a Network Slice Manager is responsible for dynamically creating and destroying instances of a network slice and map and load them to available physical host platforms (e.g. routers, switches, servers). These host platforms will be deployed across the operator network possibly across separate and distant geographic locations. The Network Slice Manager, as well as the target host platforms cannot be trusted. How does the Network Slice Manager know that the host platform on which a network slice is to be run is an operator authorized platform? Also, how do the host platforms know that the Network Slice Manager with which they are interacting has been granted appropriate authority by the operator?

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.

2.2.2 Recommendation

Network Slice Managers and host platforms should support mutual authentication. Specifically, Network Slice Managers should authenticate the host platforms in the networks they control on which they want to load network slice instances. Hosts platforms should also authenticate the Network Slice Manager before allowing the instance of the network slice to be loaded.

2.3 Key Issue 3: Impersonation attacks against a Network Slice instance within an operator network

2.3.1 Description

A Network Slice Manager could support provisioning functionality for already deployed and running network slice instances. Imagine the need to add subscriptions to an already deployed network slice instance. How can the Network Slice Manager guarantee that the correct and authorized instance of a given network slice is being provisioned? This problem is made more difficult to address due to the possibility that virtual elements that run within the slice might have moved and/or might have been destroyed and replaced by a newly created instance of an equivalent element type. This could have happened for non-malicious purposes, maybe due to failure of the element itself and the need to restore lost functionality with a new instance of the failed element. It is difficult to distinguish a scenario driven by a malicious actor and one driven by need to support failure conditions.

Impersonation attacks against a network slice instance impact all services supported by that network slice instance and would allow the corruption, removal, disclosure and interruption threats described in the previous section

2.3.2 Recommendation

All virtual functions to be contained within a Network Slice instance need to be authenticated and their integrity verified.

2.4 Key issue 4: Impersonation attacks against different Network Slice Managers within an operator network

2.4.1 Description

Consider the need to create a network slice instance that spans multiple physical networks within an operator network. Assume that these physical networks (say physical network A and B) would utilize different network slice managers (possibly because of a merger with another operator network). These network slice instances would be loaded to run on the host platforms controlled by each network slice manager that runs on each network. In this scenario the manager in, say physical network “A, would need to create a portion of the network slice instance to be loaded on host platforms deployed in the other physical network “B”. The network slice manager for physical network “B” would need to request and receive permission to accomplish this task (loading a network slice instance on a host platform). How can the network slice manager for physical network “A” trust that his negotiations are not taking place with an impostor manager?

Impersonation attacks against a network slice manager impact all services supported by that manager via the network slice instances it controls. Such attacks would allow the corruption, removal, disclosure and interruption threats described in the previous section

2.4.2 Recommendation

Network slice managers within an operator network need to mutually authenticate each other (directly or transitively) before any negotiation among them can be carried out.

2.5 Key Issue 5: Different security protocols or policies in different slices

2.5.1 Description

If different slices are offering different services, then those services may have different performance constraints, and/or different security requirements. For instance:

- The service in one slice may require extremely low latency, which constrains the security protocol in some way (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 in some other way (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).

The fact that some aspect of security is constrained in one slice shouldn't mean that it has to be similarly constrained in all slices. It is natural, therefore, to expect that security mechanisms will vary somewhat between slices – different "tuning" of security protocols (such as frequency of re-authentication), or possibly even different protocols.

However, where security varies between slices, we need to consider how well those slices are isolated from each other. If someone can attack the "lower security slice", can they then also impact the "higher security slice"? We also need to consider the security of the network as a whole: if someone can attack a "lower security slice", can they impact the whole network?

2.5.2 Recommendation

Demand a baseline security level in all slices.
The idea of having different security in different slices should never be taken as justification for having a slice with security that is too low. This means that:

- in standards, a security protocol should not be accepted for one slice if it would be considered unacceptably weak for 5G as a whole;
- in network configuration, security parameters (such as authentication frequency) should not be set lower in any one slice than would be considered acceptably secure for the network as a whole.

Adequate isolation of slices with different security levels.
The idea of working within constraints only in slices that need them, rather than forcing the resulting security compromises on all slices, is something that we encourage in principle. However, as a loosely defined guideline, where different slices have different security levels, the isolation between those slices should be at least as strong as the difference between security levels. In other words, it should always be more feasible for an attacker to target the "higher security slice" directly rather than trying to attack it via the "lower security slice".

Separate authentication of a UE accessing multiple slices at once.
If an UE can access multiple slices simultaneously, and those slices have different security levels with respect to network access, then the operator policy should request the UE to (re)authenticate separately for each slice. Otherwise the UE may authenticate itself to the "lower security slice" and thus be allowed access to the "higher security slice".

2.6 Key Issue 6: Denial of service to other slices

2.6.1 Description

By exhausting resources in one slice, an attacker may exhaust resources common to multiple slices, and hence cause denial of service or service degradation in other slices too.

The same concern applies if resources in the first slice are exhausted not by a deliberate DoS attack, but just by accident or chance (e.g. the "flash events" mentioned in [2]).

Note: when we talk about "resources common to multiple slices", we mean:

- either hardware-level resources (memory, processing power etc.) allocated to virtualized entities on the same host platform,

- or network functions providing services to multiple slices (e.g. a single HSS providing authentication vectors to MMEs in multiple slices)

2.6.2 Recommendation

When designing the systems that allocate network resources to individual slices, consider:

- At least, capping resources for individual slices, so that each such slice is allowed at most a prescribed maximum level of resource.
- Optionally ring-fencing resources for individual slices, so that each such slice is guaranteed a prescribed minimum level of resource

2.7 Key Issue 7: Exhaustion of security resources in other slices

2.7.1 Description

Suppose that an attacker wishes to "do something bad" in slice A. Normally, slice A would run its normal security protocols, and this would prevent the attack. But now suppose that 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.

2.7.2 Recommendation

Ring-fence network resource for security protocols, so that a slice always has the ability to run them, no matter how heavy the resource exhaustion in other slices. (Note: the recommendation in Key Issue 6 talks about ring-fencing resources in a more general sense, but here we are specifically talking about ring-fencing resources available to run security protocols.)

2.8 Key Issue 8: Side channel attacks across slices

2.8.1 Description

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. Examples might be observing the power consumed when the code runs, or observing the time that the code takes to run – possibly while influencing other inputs or running other code so as to influence the contents of the cache, or while somehow inducing faults on the platform.

Suppose that slices A and B share some underlying hardware. If an attacker can observe or influence how code runs in functions in slice A, she may be able to affect the running of code in functions in the slice B machine, or (more importantly) 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.

If the two slices run services with a similar level of sensitivity, then this doesn't create a significant benefit for the attacker. But if slice B's service is more sensitive, this could provide an effective, indirect way to attack it.

2.8.2 Recommendation

There are two main complementary recommendations to address this threat.

The first is simply about the strength of isolation of virtual machines: observing or influencing how code runs in one virtual machine should not allow an attacker to influence or deduce anything about how code runs in another virtual machine on the same hardware. (This is not specific to network slicing.)

The second is to avoid co-hosting on the same hardware slices that have very different levels of sensitivity, or very different levels of vulnerability to influence by an attacker. For example, avoid co-hosting one slice that supports a particularly sensitive service and another slice that supports application layer code being run on the same hardware (e.g. Mobile Edge Computing). Similarly, avoid co-hosting one slice that supports a particularly sensitive service and another slice that uses new, experimental network code. In both cases we make this recommendation because there is a greater risk that the "other slice" will have exploitable bugs that could lead to hostile code running in the slice A machine.

The two recommendations are complementary alternatives. If the first recommendation can be met very robustly, then there is not much residual value in the second; but if the first may not be robustly satisfied, the second has value.

A third, much more general recommendations is: consider the threat of side channel attacks across slices in any relevant certification or security assessment scheme.

2.9 Key issue 9: Hybrid deployment models

2.9.1 Description

We often talk about Network slicing and how it will be prevailing in 5G, but we have also to consider (at least for the beginning of 5G deployments) that some network functions might not be virtualized (e.g. HSS/AuC). This could be because of operator preferences, or because, during the first phases of 5G deployment, not all network functions may be available in virtualized mode, constraining the operators to deploy a mix of regular functions and virtualized ones.

2.9.2 Recommendation

The 5G architecture should ensure that such deployments maintain the same (or better) 5G security

2.10 Key issue 10: Sealing between slices when the UE is attached to several slices

2.10.1 Description

The key issues 5 and 6 describe sealing between slices at the network level, but a UE could be attached to several slices, which 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 between 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 to when a laptop can access open Internet on the one hand, and on the other hand it may access an enterprise network. Then, sensitive enterprise data may leak to the open Internet 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 notion.

2.10.2 Recommendation

- If a UE can consume services from multiple slices, then the impact of this on Network Slicing security needs to be studied.
- Security mechanisms to address this should exist in the network and could also potentially exist in the UE.

Note: The ME (Mobile Equipment) is typically considered as untrusted to the network.

Note 2 : In 3GPP terminology (cf TR 21.905), “Mobile Equipment” (ME) means the end user device *excluding* the UICC domain, while “User Equipment” (UE) means the combination of Mobile Equipment and UICC domain.’