ngmn
the engine of broadband
wireless innovation

5G security – Package 3: Mobile Edge Computing / Low Latency / Consistent User Experience

5G security – Package 3: Mobile Edge Computing / Low Latency / Consistent User Experience

by NGMN Alliance

Version:	**2.0**
Date:	**20 February 2018**
Document Type:	**Final Deliverable (approved)**
Confidentiality Class:	**P - Public**
Authorised Recipients: (for CR documents only)	

Project:	**NGMN – 5G Security**
Editor / Submitter:	**Rémy HAREL (ORANGE) / Steve BABBAGE (VODAFONE)**
Contributors:	**NGMN 5G security group**
Approved by / Date:	**NGMN Board, 6th February 2018**

The information contained in this document represents the current view held by NGMN Ltd. on the issues discussed as of the date of publication. This document is provided "as is" with no warranties whatsoever including any warranty of merchantability, non-infringement, or fitness for any particular purpose. All liability (including liability for infringement of any property rights) relating to the use of information in this document is disclaimed. No license, express or implied, to any intellectual property rights are granted herein. This document is distributed for informational purposes only and is subject to change without notice. Readers should not design products based on this document.

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 ETSI MEC) 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 focused on security threats or flaws that could emerge through the Network Slicing use in 5G. The present document focuses on Mobile Edge Computing, Low Latency and Consistent User Experience. Mobile Edge Computing (which is part of the slightly broader concept of Multi-Access Edge Computing) and Low Latency allow new types of services. The NGMN 5G SEC group studied the security threats, frauds and vulnerabilities that such concepts could introduce in 5G and provide security recommendations to mitigate them.

An important takeaway from this document is a note of caution on the very low latency targets often stated for 5G. Some of these targets seem to be of questionable value – and if latency targets are too aggressive, this can pose quite severe constraints on security mechanisms. The benefits and implications of very low latency need to be carefully weighed against each other; moreover, operators should not compromise generic 5G security in order to reach very low latencies that are only required for specific use cases. Where use case requirements would necessitate security compromises, those compromises need to be weighed up against the benefits of very low latency for that specific use case in order to derive a more precise and balanced set of requirements including security considerations. Since the release of the NGMN 5G White paper, new studies and experiments are available, making the 5G environment richer than ever. Such additional material should be used to start a new study (NGMN) to help refine the latencies needed by the 5G low-latency use cases. A publication with refined latencies for the main low latency use cases would be very useful to 5G actors.

Document History

Date	Version	Author	Changes
June, 27, 2016	V 0.1	Remy HAREL (ORANGE)	Creation of the document
June, 30, 2016	V 0.2	Remy HAREL (ORANGE)	Addition of new Key Issues
July, 01, 2016	V 0.3	Steve BABBAGE (VODAFONE)	Minor corrections
July, 01, 2016	V 1.0	Remy HAREL (ORANGE)	Final draft edition
July, 08, 2016	V 1.1	Steve BABBAGE (VODAFONE)	Extension to the abstract
Sept, 09, 2016	V1.2	Steve BABBAGE (VODAFONE)	Revision after Board feedback
Oct,25, 2016	V1.3	Remy HAREL(ORANGE)	Revision after 2nd Board feedback
Aug, 16, 2017	V1.4	Steve BABBAGE (VODAFONE)	Update after liaison from ETSI LI
Aug, 28, 2017	V1.4.1	Steve BABBAGE (VODAFONE)	Update after SCT feedback
Sep, 18, 2017	V1.4.2	Steve BABBAGE (VODAFONE)	Update after SCT feedback
Oct, 04, 2017	V1.4.3	Steve BABBAGE (VODAFONE)	Update after SCT feedback
Feb, 20, 2018	V2.0	Steve BABBAGE (VODAFONE)	Very minor update after Board feedback

REFERENCES

Ref	Document
1	NGMN 5G White Paper, version 1.0, Feb. 2015.
2	ETSI Mobile-Edge Computing – Introductory Technical White Paper, Issue 1, Sep 2014.

GLOSSARY

Abbreviation	Meaning
3GPP	3rd Generation Partnership Project
AKA	Authentication and Key Agreement
AuC	Authentication Server
CDN	Content Delivery Network
CN	Core Network
DNS	Domain Name Service
DoS	Denial of Service
DDoS	Distributed Denial of Service
DSS	Distributed Subscriber Server
E2E	End to End
EPS AKA	Evolved Packet System Authentication and Key Agreement
ETSI	European Telecommunications Standards Institute
HSS	Home Subscriber Server
HTTP	Hypertext Transport Protocol
HTTPS	Hypertext Transport Protocol Secure
IP	Internet Protocol
LEA	Law Enforcement Agency
LI	Lawful Interception
LTE	Long Term Evolution
MAC	Message Authentication Code
ME	Mobile Equipment (from 3GPP nomenclature)
MEC	Mobile Edge Computing (or Multi-Access Edge Computing)
PDGW	Packet Data Gateway
QoS	Quality of Service
RAN	Radio Access Network
RD	Retained Data
SA	Security Association
SDN	Software Defined Network
SGSN	Serving GPRS Support Node
UE	User Equipment (from 3GPP nomenclature)
UICC	Universal Integrated Circuit Card
VNF	Virtualized Network Function
VPN	Virtual Private Network
Wi-Fi	Wireless Fidelity

Contents

1 MOBILE EDGE COMPUTING

1.1 Key Issue 1: Billing Risks from MEC deployments

1.1.1 Description

In conventional cellular networks, billable traffic is routed into the core network (e.g. SGSN, PDGW). 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 of what billable types etc.) and so helps prevent billing errors, or deliberate fraud.

However, in MEC, significant 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). In particular, 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 disputed bills with the subscriber), and of deliberate billing fraud. Under-billing is an obvious 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), but so is over-billing (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.

1.1.2 Recommendation

MEC might be provided on an “all you can eat” model, but this is not a very flexible solution. Alternatively, the core network could benefit from periodically polling the UE, and asking it how much data (how many packets or bytes, what billable types, service identity etc.) it has transmitted or received recently. This would provide a cross-check on any charging records received from the edge, and so detect various types of billing error or fraud. Another possibility is to use a push mechanism, where the UE sends reports based on a predefined policy or schedule.

There is an existing 3GPP mechanism “Signalling procedure for periodic local authentication” (3GPP TS 33.102, clause 6.4.7, and 3GPP TS 33.401, clause 7.5). That mechanism runs between the UE and visited network. A similar or new mechanism could be developed in 5G, ideally with secure signaling between the UE and the home network.

The appropriate polling/reporting frequency might vary; for instance, a UE sending occasional small readings might be polled (or report) less frequently than another UE that consumes data much faster.

Fraud which involves both a hacked edge **and** a hacked UE may still be undetectable, unfortunately.

1.2 Key Issue 2: Third party applications on the same platform as network functions

1.2.1 Description

A likely (although not essential) model for Mobile Edge Computing is that edge computing applications will run on the same physical platforms as some network function. These will be third party applications, not controlled by the operator 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 thing themselves.

1.2.2 Recommendation

One approach that could be imagined would be to create some sort of application quality assurance framework, so that only trusted applications are allowed to run. This could be expensive and restrictive so another approach might be to design the platform in such a way that the mobile operator doesn't need to trust the edge computing applications. This will typically be achieved by running both the edge computing applications and the network function(s) in robustly segregated virtual machines, providing an assurance of confidentiality for sensitive data/information between VMs running on the same physical platform, and between a hypervisor and the host OS. (Examples of sensitive data/information would include data relating to lawful interception, retained data, or information that could compromise the confidentiality of such sensitive functions.)

Based on mobile operator's policy the solution should ensure that network functions and potentially some applications can be given a higher priority than others.

In addition, mobile operators should (contractually or legally) avoid putting themselves in a position where they are held responsible if third party edge computing applications go wrong.

1.3 Key Issue 3: Third party applications allowed to influence the network

1.3.1 Description

One of the use cases identified in ETSI's Introductory Technical White Paper on Mobile Edge Computing [2] is "Application-aware Performance Optimization" – the idea that "Application-aware cell performance optimization for each device in real time can improve network efficiency and customer experience". To some extent, this means that information provided by the application can have an influence on how the RAN is configured. If this influence is too great, it could cause severe degradation or denial of service to other users. Some applications might starve competitor applications (and their customers) of radio resource, either accidentally or maliciously.

1.3.2 Recommendation

When designing the mechanisms by which applications can influence network configuration, it must not be assumed that all such applications are trusted/well-behaved. On the contrary, the design should be reviewed on the assumption that at least some applications are malicious and will distort network performance (e.g. quality of service) as much as they are able … and operators must assure themselves that that distortion will still be within acceptable bounds.

1.4 Key Issue 4: Services to third party MEC application providers

1.4.1 Description

Where mobile operators manage MEC application servers that host third party applications, there is an opportunity for operators to provide security / assurance services for those applications.

1.4.2 Recommendation

Consider performing integrity assurance checks on applications at installation or upgrade, or after a server restart.

Consider exposing security services APIs to sufficiently trusted third party MEC applications, e.g. for user identification.

1.5 Key Issue 5: User Plane Attacks in a Mobile Edge Computing Environment

1.5.1 Description

In 5G, a likely architecture would move cellular functionality to the edge of the network and in so doing would also move the IP connectivity layer closer to the user. In this new architecture server computation and content would move closer to the edge of the network, at the macro cell level, with server level computation and portions of content stored in caches provided by the cellular operator. It is likely that the current functionality of DNS resolution and content delivery networks would also move closer to the edge of the network. In this new architecture IP connectivity would terminate at the edge (macro cell level) of the operator network. This situation will alleviate many challenges faced when optimizing encrypted video content encrypted 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 namely:

1. Security Threats that target the content server using protocols like HTTP/HTTPS
2. Security Threats against content caches (cache poisoning attacks)

Besides traditional attacks against servers and caches (e.g. via HTTP response splitting), new type of attacks will be possible. For instance, since a very large number of caches at the edge of the network would be deployed in this new architecture with possibly considerable smaller capacity than done today (where operator/CDN caches tend to be deployed closer to the exit/entry to the internet to serve larger population of users), attackers will be able to easily overwhelm these caches with request for content not likely to be used by non-malicious users. This situation would result in filling local caches with “useless” content unusable by subscribers and would, for all practical purposes, have the effect of disabling these caches. Such attacks can cause major disruption to the latencies service level agreements committed to by operators.

Although not really a user plane attack, another challenge to consider when deploying multiple caches is the correct and timely delivery (and revocation) of website certificates to those caches.

1.5.2 Recommendation:

It is important to understand the security implications (new vulnerabilities and threats) of deploying a large number of caches in operator networks. It is also critical to understand the implications of possibly serving content to users tied to the same content-session (e.g. video session) from multiple caches as users move and are potentially served by multiple caches. It is also important to clearly recognize the security implications of delegating execution of end servers’ code to computing resources deployed at the edge of the network. Research in these areas will likely help our understanding of the new security challenges that exist with this new proposed 5G architecture.

1.6 Key Issue 6: Storage of Sensitive Security Assets at the Edge

1.6.1 Description

If sensitive security assets are compromised at virtualised functions at the edge, an attacker could maliciously reuse them to gain connectivity or carry out a spoofing, eavesdropping or data manipulation attack. The attack surface increases the more functions there are (and the more local data centres hosting them are involved).

1.6.2 Recommendation

Sensitive Security Assets stored at the mobile edge should be encrypted so that these assets are not compromised. The security level of the storage should ideally be assured according to a recognised scheme.

If the sensitive assets are encrypted at the edge, then the decryption keys will also need to be available there, and so they in turn need to be protected. Furthermore, the threat to sensitive assets while temporarily decrypted also needs to be addressed.

1.7 Key Issue 7: Exchange of Sensitive Security Assets between core and Mobile Edge

1.7.1 Description

If sensitive security assets are compromised when exchanged between the core and the edge, an attacker could maliciously reuse them to gain connectivity or carry out a spoofing, eavesdropping or data manipulation attack.

1.7.2 Recommendation

Sensitive Security Assets exchanged with the mobile edge should be encrypted so that these assets are not compromised. The security level of the entities securing the data both at the core and the edge should ideally be assured according to a recognised scheme.

1.8 Key Issue 8: Trust Establishment between functions at the core and at the edge

1.8.1 Description

Should Network functions be present at the edge, then the functions both at the core and the edge need to trust that (a) they are exchanging with the entity they expect, and (b) that entity is still doing what it is expected to do. The second point is especially relevant where edge network functions are outside of the premises of the mobile network operator, and so there is an increased risk of physical tampering.

1.8.2 Recommendation

Subject to risk assessment by the operator, a mutual authentication between each physical and virtual resource which need to communicate, be they at the core or at the edge, should be carried out. This should be done upon service start/restart; likewise a new mutual authentication procedure should be carried out following the upgrade of any function.

The security level of the entities securing the mutual authentication framework both at the core and the edge should ideally be assured according to a recognised scheme. This assurance should include an integrity check of the network function code, to compare it with that of the original blueprint installed; this is particularly recommended in the case of premises not controlled by network operators.

1.9 Key Issue 9: Security of Communications with the MEC Orchestrator

1.9.1 Description

As the physical connection from the core to the edge is outside of the premises of the mobile network operator, the physical or virtual links are open to compromise and instructions transmitted by a MEC orchestrator in the core to the mobile edge could be compromised, spoofed or modified.

1.9.2 Recommendation

Instructions exchanged between the orchestrator and network resources at the core and the mobile edge should be integrity and confidentiality protected so that they are not compromised. The security level of the entities at the core and the edge securing the exchanges should ideally be assured according to a recognised scheme .

1.10 Key Issue 10: Law enforcement requirements for MEC deployments

1.10.1 Description

Operators are required to provide LEA Support including Lawful interception (LI) and Retained Data (RD) capabilities for traffic carried on their networks; typically this functionality is supported at nodes within the core network.

However, traffic carried from the UE to an application at the network edge is currently designed to avoid the core, and hence would avoid the usual intercept points.

Moreover, in the context of MEC placing multiple additional LI points around the network edge raises security risks:

- there will be many more LI points than in traditional deployments
- and also edge nodes are likely to be more exposed to attack than core nodes.

1.10.2 Recommendation

In the context of MEC, it is recommended that LI and RD collection functions are implemented at the edge of the network, alongside or as part of the functionality being intercepted. Any edge node including LI / RD collection features must support strong physical security requirements similar to core network sites. Further work would be required to examine specifically how and where the LI/RD functionality should be included in a network architecture. The following considerations all apply when LI functionality (and RD collection functionality) is provided at the edge rather than the core of a network. It is essential that the following sections of the present document are handled effectively:

- 1.1 Billing risks. The same considerations for billing will also need to be handled for meeting RD requirements.
- 1.2: Robustly segregated virtual machines, providing an assurance of confidentiality for sensitive data/information between VMs running on the same physical platform, are essential for the reasons described in section 1.2 and also for LI and RD requirements. Section 1.2 puts forward some reasons for avoiding the "quality assurance framework" model; the LI and RD requirements also support the reasons for avoiding the “quality assurance frameworks” described in section 1.2.
- 1.4 Assurance checks at installation or upgrade will be essential technology from an LI/RD point of view, in terms of assurance of all components to avoid malware being installed in places where it would be able to attack LI/RD functions, and specifically in installing or upgrading LI functionality.
- 1.5 User Plane attacks. This consideration is very relevant, to prevent attack on LI/RD functions.
- 1.6 Storage of sensitive assets at edge. This is very relevant for storage of LI target identifiers.
- 1.7, 1.8 and 1.9 (Exchange of sensitive security assets between core and mobile edge, Trust Establishment between functions at the core and at the edge, Security of Communications with the MEC Orchestrator). These will all be critical for setting up LI functionality at the edge and for updating new target identifiers, also for secure delivery of LI traffic.

2 LOW LATENCY

5G description is usually defined with many examples of new use cases, and some of them are presented to be in a need of low latencies (5-10ms), or very low latencies sometimes (1-2ms). The NGMN white paper describes a few of these use cases, such as remote surgery, automated driving (transport), tactile internet, video gaming, pervasive video….

The concept of low latencies may require specific network architectures or deployments, which could introduce new security threats or problems. The NGMN 5G security group focused on the impact of the low latency concept on 5G security.

2.1 Key Issue 1: Latency targets for security mechanisms should not be too aggressive

2.1.1 Description

Overly aggressive latency targets for 5G (e.g. <10ms or <1ms latency) may compromise system security, or else entail a completely new security architecture. Very great care needs to be exercised before accepting security compromises, or expensive security changes – and this includes a clear assessment of whether / how strongly the latency requirements are truly justified by the envisaged use cases.

3GPP security mechanisms include authentication and key agreement, with periodic re-authentication (which requires round trips to the home network), signalling to the core network to manage security associations and session key updates, secure handovers between cells, and basic cryptographic operations of encryption and decryption, creation and verification of MACs etc.

Authentication to the home network becomes extremely difficult within a round trip of <10 ms (for speed of light reasons, the home network can be at most 1500 km away), and core network signalling is very challenging within a round trip of <1 ms (core network nodes can be at most 150 km away). It is important to note that currently we don’t have any business use case or service requirement for such low latency needs regarding the authentication or control plane between the UE and the core network.

Even basic crypto-operations become a challenge with <1 ms round trip time. If say 10% of that latency is consumed by the crypto, this requires at most 25 micro-seconds for each send and receive operation, or at most 12.5 micro-seconds for each crypto operation; compare that with http://csrc.nist.gov/groups/ST/lwc-workshop2015/presentations/session7-vincent.pdf.

In any case, before accepting such extreme changes or compromises to security, it is essential to assess clearly how strong those use cases are, and whether they really do require such low latencies. NGMN has agreed to initiate a comprehensive study of low latency use cases, to clarify what latency they really need, in precise terms (latency from where to where; one-way or round trip; latency only on the user plane, or also of control plane events; etc).

2.1.2 Recommendation

For most anticipated 5G services, a target latency of **30-50ms** for security mechanisms seems acceptable. Such a target can probably be achieved with relatively minor changes to conventional 3GPP security architecture. In most cases, this would also be compatible with <10ms latencies on the user plane traffic itself e.g. once keys are established, basic crypto-operations can all be completed in <<10 ms.

Some niche services (e.g. ones involving a high level of automation and super-human reaction times) may need a special network slice with a custom security architecture, where there is a trade-off between security and latency.

Timing of crypto-operations itself for the user plane data becomes a problem, but can be mitigated by performing most of the heavy crypto pre-emptively (e.g. in idle mode) and perhaps by moving the real-time encryption/integrity operations to a lower layer of the stack. For some services, it may be acceptable just to encrypt with a fast stream cipher. Based on the operator's decision, security of user plane data may need to be dropped entirely for some ultra-low latency services, with only intermittent integrity checks conducted via the signalling plane (for example, to keep track of how many packets/bytes have been transmitted or received by the UE).

In any case, before accepting such extreme changes or compromises to security, it is essential to assess clearly whether the intended use cases really do require such low latencies.

2.2 Key Issue 2: Subscriber authentication within the visited network

2.2.1 Description

Consistently low latency may – although this is by no means confirmed – entail very fast authentication procedures at attachment or handover. This may force subscriber authentication to be done entirely within the visited network, as it is simply not possible to perform a round trip to the home network. Observe that even a 50ms latency target would prevent a round trip to a home network more than 7500km away (so would be a barrier to roaming across continents).

2.2.2 Recommendation

It may be possible to combine low latency on the user plane with high latency on the signalling plane. User plane latency can be minimized by re-using an old security association (SA), while in the meantime running AKA and acquiring a new security association. However, this would still impose a high latency at initial attachment to a network (before the first SA is established), and would require persistent caching of old SAs by both the UE and visited network, so weakening security (there is more risk of an old key leaking and being abused). Further, if either the UE or visited network node has purged the old SA, the user plane will have to wait while a new SA is established. This may be unacceptable for some use cases.

If low latency is also required on the signalling plane, **then** the solution of "data efficient re-keying" discussed in TR 33.860 (Annex B2) could allow for *low latency* re-keying of a security association, based on an intermediate key K_{med} held by both the visited network and the UE (e.g. within the UICC or other secure element).

Another low latency solution could be to delegate some of the functions of the HSS to the visited network. We might call it a Delegated Subscriber Server (DSS). There will be only one "real" HSS, but there could be multiple DSSs. Using the GSM term "Ki" for the long term subscriber secret key:

- Ki remains only in the HSS.
- Each DSS needing to authenticate this subscriber receives a key Ki', derived from Ki in a one-way fashion. No DSS should be able to work out the actual Ki, nor the Ki' that's provided to any other DSS.
- The authentication and key agreement algorithm takes Ki' as input. The UICC (or its 5G equivalent) also has the necessary information to derive Ki'.
- Ki' could be derived just once for any DSS, or could be refreshed periodically.

There are however disadvantages with this approach since it places authentication and related billing controls in the visited network, creating a risk (e.g. fraud) for the home operator. It is important to note that the home operator cannot control how securely long-term keys are stored in the visited network. There is also a need to standardize the authentication and key agreement algorithms used between the UE and the DSS.

Most of these solutions would still require a round-trip to the core of the visited network, so would be incompatible with ultra-low latencies (<1 ms). Alternatively, the visited network could replicate its own core security functions around the edge, which is likely to be very expensive, and would compromise security, as there is much more risk of some of these functions being compromised.

Any solution needs to comply with local laws/regulations. It is typically a requirement that LI can take place for roamers in the visited network without relying on the cooperation of the home network. Moreover, the identities of LI targets must not be visible outside the country in which LI is taking place (so in the roaming case, they must not be visible to the home network). These requirements need to be satisfied by any authentication solution.

2.3 Key Issue 3: Loss of Service on user plane during critical communication following network re-authentication request

2.3.1 Description

Existing 3GPP user plane encryption mechanisms derive the encryption keys from the control plane authentication mechanism. Should the derivation of user plane keys for encryption be bound to the control plane, it could cause a brief lapse / loss of service should a re-authentication request be triggered during a critical communications session.

2.3.2 Recommendation

No critical path should be present on the user plane. The derivation of keys for user plane encryption should not be strictly bound to the control plane as it could cause a service lapse should a re-authentication be triggered during a critical communications session. The authentication mechanism on the control plane should not impact the latency for critical communications taking place on the user plane. Other techniques need to be explored.

Note that LTE already provides decoupling of AKA authentication and change of keys on the radio interface by the means of mechanisms called “key change on the fly” and a similar mechanism would be desirable in 5G for all types of access networks.

3 CONSISTENT USER EXPERIENCE

5G will operate in a highly heterogeneous environment, characterized by the existence of multiple types of access technologies, multi-layer networks, multiple types of devices, multiple types of user interactions, etc. In such an environment, there is a fundamental need for 5G to achieve seamless and consistent user experience across time and space.

3.1 Key Issue 1: Secure storage of credentials to access IMS network

3.1.1 Description

Credentials stored in a connected device can be open to compromise or cloning of the IMS network access credentials. Malware could be deployed on a massive scale to collect these credentials should a device become untrustworthy (e.g. rooting or jailbreaking).

3.1.2 Recommendation

If separate credentials are used to access IMS networks, it is recommended to secure the access to those IMS networks credentials alongside those of the 3GPP network access credentials and to apply the same level of security to their storage as the credentials of the 3GPP network.

3.2 Key Issue 2: Access to the 5G core over non-3GPP network accesses

3.2.1 Description

We expect that 5G will encompass access to the mobile operator core, and/or to mobile operator services, via non-3GPP radio access technologies. These may be non-3GPP access networks deployed by the operator, or by a party with which the operator has a business relationship or not.

It is clear, that in some of these possible cases the fact that the subscriber has (somehow) authenticated to the non-3GPP access network will not, on its own, provide enough assurance for the mobile operator to grant access to the 5G core or services.

3.2.2 Recommendation

The access to the 5G core via non-3GPP network access should be secured by a mutual authentication and key agreement between the UE and the core network using 3GPP credentials. Additionally, in case that the non-3GPP access network is considered as untrusted, a secure tunnel relying on keys derived from the authentication process should be set up between the UE and the core.

5G User Equipment therefore needs to support secure tunnels (e.g. VPNs) authenticated using operator-derived credentials.

In 3GPP systems today, if permitted by local law/regulation, networks can use their knowledge of device models to inform security and fraud controls, as well as optimising performance and services. The continued ability to do this, even for devices that are only ever expected to connect over non-3GPP access network, would be advantageous (still subject to law/regulation).

3.3 Key Issue 3: User plane data security over less trusted 3GPP network accesses

3.3.1 Description

In some cases, the 3GPP access network may not provide strong security for the user plane data (e.g. encryption of the radio interface in the visited network). The 5G system is expected to provide mechanism allowing the home operator to provide means to mitigate these issues while respecting the national laws of the visited network.

3.3.2 Recommendation

If the visited network doesn't provide strong security on the user plane (e.g. encryption of the radio interface in the visited network), the home operator should be able to establish a secure channel for user plane data between the UE and its core network. This secure channel should be at least integrity protected; whether encryption is also allowed should be under the control of the visited network (in compliance to national laws – see in particular section 1.10). It may be that techniques such as Middlebox Security Protocols can help in allowing encryption to be enabled while still preserving the ability for national regulations to be met (work in progress in ETSI TC CYBER; see also http://conferences.sigcomm.org/sigcomm/2015/pdf/papers/p199.pdf).

5G User Equipment therefore needs to support secure tunnels (e.g. VPNs) authenticated using operator-derived credentials.

3.4 Key Issue 4: Management of credentials to access non-3GPP network accesses

3.4.1 Description

This key issue is about the management of credentials used to secure the interface between non-3GPP radio access points and the UE only, and not about the access to 5G core and services.

We expect that 5G will encompass access to the mobile operator core via non-3GPP radio access technologies. These may be access networks deployed by the operator, or by a party with which the operator has a business relationship. It is not yet clear:

- whether 5G will also include access to the operator core via radio access networks that are completely independent of the mobile operator;
- or whether 5G will include cases where access to radio networks is based on non-3GPP credentials.

Use of multiple access technologies may require multiple types of access network credentials to authenticate a user. For example: passwords for various WiFi access points; client certificates for others.

Achieving a "seamless" user experience means that the user is able to authenticate to any of the access technologies without cumbersome interaction (e.g. user is not repeatedly prompted to select credentials, enter passwords etc.) Given that the credentials are likely to be so diverse, and may not be under the control of the user's home network, these requirements are challenging.

Any network access credentials stored in a connected device could be open to compromise or cloning. Malware could be deployed on a massive scale to collect these credentials should a device become untrustworthy (e.g. rooting or jailbreaking).

3.4.2 Recommendation

UEs and 3GPP AAA server should have a general capability to derive access credentials of different types, based on an underlying 3GPP subscription credentials, and to manage such derived credentials at least as securely as keys generated by EPS AKA.

Any such derivation process should preferably provide strong key separation and forward secrecy, to ensure that a later compromise of one credential does not affect the others. The system should rely on the 3GPP credentials to derive credentials to be used to access the non-3GPP access network. These derived credentials may be of various types, symmetric or asymmetric ones. Credential derivation could be based on a long term key (like K in LTE) and/or a session key (like “Kasme” in LTE).