TRUSTED CONNECTIVITY ALLIANCE

5G Security - Making the Right Choice to Match Your Needs

Published by simalliance now Trusted Connectivity Alliance

October 2016

Table of Contents

1. Executive summary

One of the clearest messages from 'An analysis of the security needs of the 5G market', the marketing paper that SIMalliance published in early 2016, is that security in 5G is use case dependent. It was that paper and this conclusion that inspired the work of this technical paper: to further examine use cases across 5G in order to highlight this broad range of security requirements and attempt to identify effective security measures and hence an appropriate technical solution that could meet those security requirements.

The conclusions of this further work are both clear and consistent with those of the earlier document - with standardisation work at an early stage, it is vital that appropriate and robust security is built in from the outset, as this is more effective than attempting to retrofit the right level of security later. Any such security must protect subscribers, devices and their communications and also the integrity of the network itself.

For many of the use cases identified, the most secure and cost effective way to achieve this is through the use of the eUICC as a hardware based, tamper-proof repository for storing algorithms, credentials and keys. However, the paper clearly makes the case that each use case must be assessed according to its security needs and the value of the data involved, and that there are use cases where other solutions can be appropriate.

The case for this conclusion comes from a clear argument built up over the body of the technical paper:

- In a 5G future, mobile operators face a surge in data combined with a decrease in average revenue per connection (ARPC). As a result, they are looking to cut costs. The solution increasingly appears to be network function virtualisation and network slicing, both for cost and technical reasons.
- Further significant factors in 5G will include higher speeds/lower latency combined with power efficiency needs, a wider variety of actors and device types, a greater range of threats and more use of the cloud and virtualisation. In order to avoid bottlenecks and integration difficulties, new security approaches will be needed. There will be a strong need for battery optimisation, particularly for IoT and M2M sensors.
- As a result, 5G technology is expected to be built around a "network of networks" concept involving network slicing and mobile edge computing. Mission critical elements must not be shared between network slices to avoid a compromise on one affecting others.
- 5G brings security requirements that greatly add to but do not replace those of earlier generations. It will also bring a wide range of threats and a greatly expanded attack surface. Many of these additional requirements come from moves towards virtualisation and the cloud and add to the need to increase security on the network side.
- Subscriptions to the 5G network will be protected by a network authentication application (NAA) within the device that takes care of network identification, authentication and encryption. The device identity and the identity stored in the NAA should be separate and independent from each other, as is the case in earlier generations with the storage of the IMEI and IMSI/keys in separate logical entities.
- The secure tamper-resistant entity storing the NAAs must be capable of being (and should be) audited and certified by a third-party and functionally tested against a suitable industry-agreed functional compliance suite.
- Massive IoT and critical communications in particular pose specific functional requirements that will impact security.
- Remote provisioning systems must be capable of meeting requirements for secure out of the box connectivity with zero configuration. Secure access to remote provisioning should be available at all times.
- Low power consumption may be a critical requirement in some areas of massive IoT, irrespective of the connection type. Security approaches in this segment must therefore be able to work with this requirement and hence with a possible hibernate state. New efficient algorithms, authentication policies and protocols that take into account lower power consumption should be evaluated.
- Equally, solutions must be capable of meeting requirements in critical communications, where human lives may be at risk, for ultra-low latency, high throughput and high reliability.
- In IoT in particular, devices will have both a projected lifespan of as much as 15 years and only periodic connection to the network and hence oversight and upgrade. It is therefore vital that their security is built to last. Equally, many devices will be simple and low cost but security must be proportionate to the value of the data rather than the short term bill of materials cost.

Combining these factors with an understanding of risk shows that investing in security now is an insurance policy for the future of 5G. Low investment and inadequate security now will require readjustment later as hidden costs, which aren't apparent today, appear. Such hidden costs may potentially arise if work is required to remedy attacks on high value data protected by inadequate means.

Considering this, alongside the greatly superior security offered by the eUICC, and the projected growth in threats, it is clear that the wrong decision about security today will prove a false economy in the future.

2. Introduction

In early 2016, SIMalliance launched a marketing paper outlining its approach to 5G security. In it, it concluded that while there is a clear need in 5G for low latency, low power and high reliability and that any eventual security solution must take account of those needs, there will be a very wide range of use cases with differing requirements that will need to be secured. It also stated that a compelling security concept for 5G must cover more than just device security and must provide a solid proposition for the end-to end perspective that copes with the mission-critical aspects of interoperability and with scalability challenges.

The intention of this follow up technical paper is to consider, in detail, security requirements in each of the major market sectors for 5G, with the aim of making recommendations about how this can be achieved, whilst highlighting the important trade-off between ultra-low latency requirements for certain use cases, in particular in critical communications, versus security. Security by its nature will add additional processing time and latency so it is important to defend security recommendations against such challenges.

The paper starts with an introduction to 5G and its potential. It presents a layered security model, before taking each of 5G's major market segments and reviewing use cases, security requirements and suggested security strategies and mitigations. It then compares the three major technologies in play to meet those requirements, the eUICC, the TEE and the SoftSIM, for suitability for purpose.

2.1 A 5G world

2.1.1 What is 5G?

5G is the proposed next generation of mobile wireless broadband technology. As with previous generations of mobile technology, security and privacy remain fundamental underlying requirements for mobile applications and services across devices that access wireless networks.

There is currently no defined standard for 5G and technical standardisation work is just beginning. At present, 5G is a set of aspirations around desired services that are expected to be commercially available around 2020. These services are expected to place new requirements on connectivity, flexibility, cost efficiency and performance. Some companies are working on launching 5G capable networks commercially in 2018[1].

Ovum suggests that by 2021 5G networks will be operational in 20 countries, with global subscriptions set to reach 24 million by the end of 2021[2]. Its prediction concerns the enhanced mobile broadband segment of 5G and excludes any type of narrowband technology or pre-standard 5G. It expects massive IoT and critical communications uses to roll out later than 2021.

However, in June 2016 standardisation body 3GPP[3] announced a plan for the release of Release 15, the first set of requirements for 5G (3GPP's "Next Generation System") specifications with progress expected towards the end of 2016[4]. At the same time, they announced the completion of Narrowband IoT standardisation[5] which supports rates of 150kbps using LTE. LTE requires use of an UICC.

As this work progresses, it is crucial that security is built into 5G from the outset, for as SIMalliance concluded in its ***earlier paper,*** it is far more effective to take this proactive approach than to attempt to retrofit security later.

This security should protect subscribers, devices and their communications but also the integrity of the network itself.[6]

2.1.2 5G use cases and their security

Five main segments for 5G have been defined by 3GPP: massive IoT, critical communications, enhanced mobile broadband, V2X (Vehicle to X) and network operations (which underpins the four other areas).

According to Ovum, enhanced mobile broadband is the segment that will largely count for the subscriptions it predicted over the next five years[7], with massive IoT, V2X and critical communications coming later.

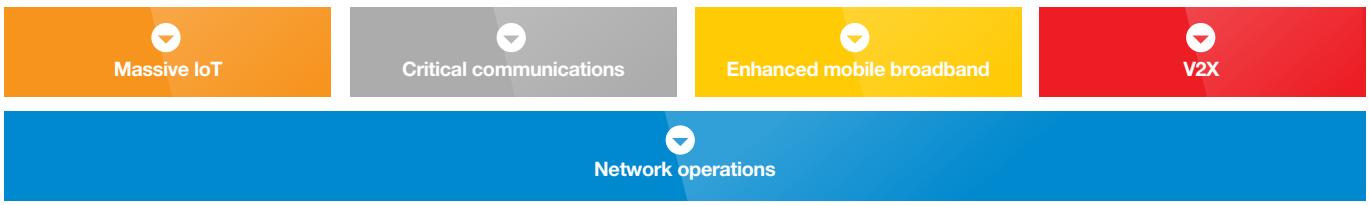

Figure 2.1: 5G segments

① http://www.ericsson.com/thecompany/press/releases/2016/01/1980613
② http://www.fiercewireless.com/europe/ovum-5g-subscriptions-to-reach-24m-by-end-2021
③ http://www.3gpp.org/
④ http://mobileeurope.co.uk/press-wire/3gpp-outlines-plan-for-5g-standardisation
⑤ http://www.3gpp.org/news-events/3gpp-news/1785-nb_iot_complete
⑥ Towards 5G Security, Horn & Schneider, Nokia, http://bit.ly/296sGDq
⑦ http://www.totaltele.com/view.aspx?ID=494217&mail=1792

Typical use cases within **massive IoT** will split between consumer and M2M and include home appliances, some wearables and machine type communications including metering, sensors and alarms. Data is likely to encompass geolocation data, sensor data, such as meter readings, and private consumer data.

Communications may be either long range, low power and infrequent or, in some cases, focused on speed. Devices may connect to the network either directly or indirectly through relay devices.

Security threats may include data manipulation, use of low cost endpoints for entry into the network, rogue devices, ransomware, equipment cloning and denial of service. There is a risk that very simple, low cost devices may not be adequately secured because of the cost that security would add to unit prices. Indeed, according to ABI Research [8], security is not front of mind for many companies building products for this sector.

In industrial IoT eavesdropping is likely to be a means to an end – for example to reverse engineer the format so you can bluff the integrity protection and carry out an instruction to the machine. In consumer IoT, the data itself might be valuable.

This exceptionally broad segment will lead to a broad range of security requirements – smart wearables for example will have different confidentiality requirements to machine sensor networks.

Key Massive IoT Threats
Data manipulation to consumer data, machine and sensor data
Unprotected endpoint used for network entry
Equipment cloning
Rogue devices
Lack of protection to low cost devices
Denial of service attack on security networks
Eavesdropping
Impersonation attacks

Figure 2.2 Threats in massive IoT

⑦ http://www.totaltele.com/view.aspx?ID=494217&mail=1792

The critical communications segment covers drones and their control, cloud driven virtual reality, smart multi-node factories, cloud driven robots, public safety, transportation and e-health.

Data is likely to include geolocation data, instructions to elements in cloud based interactions and medical, operational and situational data. Some of this data will be highly sensitive and will be communicated more frequently than in segments like IoT.

Threats may include man in the middle attacks, eavesdropping, denial of service and rogue devices. There is a risk that security may be sacrificed for speed.

Critical Communications Threats
Man in the middle attacks
Eavesdropping
Denial of service
Rogue devices
Device theft
Terrorist attacks, with device used as weapon or target

Figure 2.3 Threats in critical communications

Enhanced mobile broadband encompasses the use of tablets, mobile phones and other portable devices. It offers a range of very diverse scenarios, with indoor and outdoor use, fast moving versus slow moving devices, use on trains and planes, hotspots such as offices, crowds and high density areas, low density areas and varying levels of network capacity requirement, depending on time and other factors. Data will vary but is likely to be user data, usage data and geolocation data.

The major requirements for managing connectivity will be convenience and instant access to high bandwidth services, so again there is a risk that security may be sacrificed for speed.

Breaches may occur at both a service level (e.g. gaining a user's account credentials to log in to a service) or at a network access level (e.g. gaining a user's network connectivity credentials or piggybacking on a user's connection without them knowing). Threats are likely to be similar to those for critical communications.

5G is also expected to enable what is referred to as the Tactile Internet[9], where in addition to sight and sound, the feeling of touch is also possible over the internet.

This will make more services possible such as remote surgery and industrial automation. It will also enhance online shopping and facilitate multiple personal online activities and therefore is relevant to both critical communications and enhanced mobile broadband. This will require extremely low latency in combination with high availability, reliability and security[10].

Enhanced Mobile Broadband Threats
Man in the middle attacks
Eavesdropping
Data manipulation
Rogue devices
Device theft
Content piracy

Figure 2.4 Threats in enhanced mobile broadband

9 https://eandt.theiet.org/content/articles/2015/03/tactile-internet-5g-and-the-cloud-on-steroids/
10 http://www.itu.int/en/ITU-T/techwatch/Pages/tactile-internet.aspx

V2X covers vehicle to vehicle, vehicle to infrastructure and vehicle to anything communications. It combines elements of enhanced mobile broadband and critical communications, in particular the latter's requirement for ultra-low latency.

Once again, data is likely to include geolocation data and instructions to elements in cloud based interactions as well as user data, usage data infotainment and even premium content data.

Threats may include attacks against the vehicle as well as misusing the vehicle itself as a threat agent against other vehicles or pedestrians.

V2X Threats
Vehicle-jacking (stealing the vehicle and instructing it using the normal means or taking remote control of the vehicle)
Vehicle-hacking, for multiple purposes
Man-in-the-middle attack (information gathering as a means to carry out other attacks listed)
Vehicle tracking
Spoofing attack (information coming from unauthorised infrastructure/other vehicle source)
Data manipulation (diagnostics such as repair conditions, fuel/oil/electricity levels, deactivating/bluffing reading from sensors…)
Piracy of in-vehicle entertainment
Terrorist attacks / ultra-violent video gaming scenarios (combination attack: deactivate pedestrian detector + vehicle-hacking/jacking)

Figure 2.5 Threats in V2X

3. Technical evolutions to date

This chapter briefly reviews how mobile telephony has developed to date, providing context for how 5G will develop. It also highlights characteristics and approaches, such as remote provisioning, that will continue to be relevant going forward.

3.1 2G - 3G - 4G

It is now 25 years since GSM or 2G digital cellular networks were first deployed (in Finland in 1991). In that time technology has developed far beyond voice only telephony to include first elementary and then more advanced data communications.

2G refers to the CDMA and GSM standards, which were succeeded by 3G UMTS and 4G LTE, both of which were developed by 3GPP.

While 1G phones were analogue, vulnerable to eavesdropping and did not feature SIMs, 2G introduced the SIM, privacy through the Temporary Mobile Subscriber Identity (TMSI) and strong authentication. 3G brought mutual authentication and stronger algorithms, introducing signalling integrity and moved encryption deeper into the network. 4G returned user data encryption to the base station and introduced more elaborate key management ⑪.

Initially these standards (other than analogue) were based around the SIM or UICC. In addition, the embedded UICC (eUICC) was first standardised by GSMA. In the future however a more diverse range of form factors is expected.

Initial reasons for the introduction of the SIM card were portability, separation between the device and the subscription, plus the need to secure the user subscription and manage confidentiality. None of these requirements have gone away and with the changes brought by the arrival of 5G, such as virtualisation and the cloud, these security needs can only increase. Furthermore, the arrival of security on the network side is also a trend that will increase with 5G, with network function virtualisation producing a need to secure the APIs and functions available on the network.

3.2 Diversity of form factors and features

To date, the removable SIM or UICC has come in a range of sizes. The trend has been towards a decrease in physical size, although functionality has been size independent.

The very first SIM card was ISO payments card format (1FF), but in 1996 this was followed by the mini-SIM (2FF), which had the same contact arrangement as 1FF. So did the micro-SIM (3FF), maintaining backward compatibility. It was developed by ETSI and other standardisation bodies and launched in 2003. The nano-SIM (4FF) appeared in early 2012.

The rise of the Internet of Things (IoT) and Machine 2 Machine (M2M) communications and the ongoing reduction in size of both modules and devices, for example sensors and meters, as well as use case scenarios requiring devices to suit more rugged environments, created a need for an embedded form factor. This can be used in a hermetically sealed device such as an industrial meter or a sealed component in a fleet management system, which cannot be opened to swap out a removable SIM. The embedded SIM takes the standard SIM contacts and makes them available on a surface mounted package, with 2 options according to the ETSI specification - MFF1, which is socketable or MFF2 which is solderable. While use of this is primarily envisioned for the M2M space, some handset manufacturers are considering the embedded SIM for direct use in handsets.

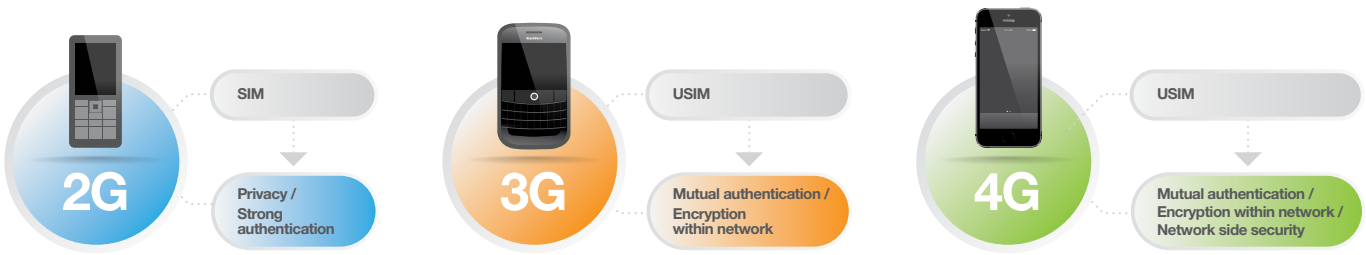

Figure 3.1 History of mobile networks

⑪ http://www.ericsson.com/lb/res/docs/whitepapers/wp-5g-security.pdf

The UICC has also evolved in terms of features over this period. Its authentication capabilities have grown from subscriber authentication to mutual authentication of subscriber and base station. Multiple features and applications have been added too, such as:

- The IP Multimedia Services Identity Module (ISIM), which identifies and authenticates the user to the IP Multimedia Subsystem, the framework for delivering IP multimedia services over mobile.
- The Generic Bootstrapping Architecture (GBA), used for key establishment between UE (User Equipment) and third party application.
- WLAN access.
- USIM Application Toolkit (USAT) application pairing.
- UICC carrier privileges as an extension to GlobalPlatform Secure Element Access Control (SEAC) [12].
- Mission Critical Push-to-Talk (MCPTT).
- Support of Isolated E-UTRAN Operation for Public Safety (IOPS)New USIM authentication type for relay node authentication.

3.3 GSMA Embedded SIM Specification

The introduction of the embedded form factor in the shape of the eSIM or eUICC created a need to manage subscriptions remotely because of the difficulty of accessing and changing these embedded secure elements.

This is covered by the GSMA Embedded SIM Specification, which provides a "single, de-facto standard mechanism for the remote provisioning and management of machine to machine (M2M) [13] connections, allowing the "over the air" provisioning of an initial operator subscription, and the subsequent change of subscription from one operator to another" [14]. The GSMA Embedded SIM Specification 'Remote Provisioning Architecture for Embedded UICC Technical Specification' [15] references SIMalliance's eUICC Profile Package: Interoperable Format Technical Specification v2.0.[16]

eUICCs, defined by SIMalliance as a "UICC which is not easily accessible or replaceable, is not intended to be removed or replaced in the terminal, and enables the secure changing of Subscriptions" [17], are widely used today in M2M deployments, where the technology is mature. The usage of eUICCs in connected consumer devices is still in its infancy.

The early signs of a consumer market for eUICCs has been driven by two key external market factors 1) regulatory activity (within the M2M space) arising from the need for a technical solution allowing service providers to remotely switch from one MNO to another in order to promote competition. and 2) OEM roadmaps, which have seen high profile consumer devices launch with eUICCs, although to date this is limited to wearables, due to their requirement for small form factors. These two factors have combined thanks to an industry desire to see technology consistent across both sectors, according to GSMA [18].

3.4 The value of eUICC with remote provisioning

eUICCs combined with remote subscription management processes provide unrivalled security to protect subscriptions in unattended devices against attacks. eUICCs offer an easy route to security certification (it is simpler to certify a small, distinct piece of hardware, with established and mature certification processes, than a larger machine or a more complex system on chip).

The remote provisioning also eases the device manufacturing and logistic (separation between the subscription and the device).

The advent of GSMA's Remote SIM Provisioning (RSP) specification has transformed the SIM from a static personalised-at-production element to something much more dynamic (in the form of an embedded UICC) which can be provisioned for service on-demand and whose lifecycle including that of the NAA - and connectivity parameters within the NAA - can be modified remotely. This could provide flexibility for a service providers evolving business relationships with various global and local network operators providing connectivity. This technology is therefore well adapted to the evolving needs of the 5G environment.

As a result standardised remote subscription management across eUICCs, regardless of their source, results in time and development efficiencies for MNOs and the wider remote provisioning ecosystem. It enables service providers to provision their fleet, or installed base of devices, more rapidly and in a unified way across diverse terminals, MNO customer management systems and eUICCs. Files are exchanged securely and provisioned back to a subscription manager, which is the entity that operators use to securely encrypt their operator credentials ready for over the air installation within the eUICC. It then securely delivers the encrypted operator credentials to the eUICC and then, once the credentials are installed, remotely manages the eUICC thereafter (enable, disable and delete the credentials as necessary during the product's lifetime) [19].

[12] https://source.android.com/devices/tech/config/uicc.html
[13] This is equally applicable to consumer devices.
[14] http://www.gsma.com/connectedliving/embedded-sim/
[15] http://www.gsma.com/connectedliving/wp-content/uploads/2012/03/SGP-02-v3-0.pdf
[16] http://simalliance.org/euicc/euicc-technical-releases/
[17] http://simalliance.org/euicc/euicc-technical-releases/
[18] https://www.gsmaintelligence.com/research/
[19] http://www.gsma.com/connectedliving/embedded-sim/how-it-works/

GSMA operates a Security Accreditation Scheme that enables operators to assess the security of their eUICC equipment suppliers and subscription management service providers [20].

The end result is simple application provisioning and lifecycle management and more scalability and flexibility within the remote provisioning ecosystem.

3.5 The trusted relationship between MNOs and SIM vendors

The traditional relationship between SIM vendors and MNOs has been based on the provision of physical SIM cards. SIM vendors use their unique expertise to personalise SIMs with subscriber profiles, to meet MNO requirements. Each profile contains very sensitive data including connectivity parameters to access the MNO's network as well as a set of unique identifiers and credentials (keys). The MNO requires extreme levels of security implemented within the SIM vendors' facilities as well as trust in the entity responsible for managing this data (the SIM vendor), a trust that has been established and proven over many years. In a future world where eUICCs are more common within connected devices, this relationship between MNO and the SIM industry will remain strong, thanks to the trust that already exists. SIM vendors will supply their services to MNOs, in order to continue building profiles and managing all sensitive MNO data within them.

3.6 Secure element technology beyond mobile network access credential protection

The SIM or UICC is a type of secure element. According to GlobalPlatform, the industry body that identifies, develops and publishes specifications which facilitate the secure and interoperable deployment and management of multiple embedded applications on secure chip technology, a secure element (SE) is a tamper-resistant, distinct hardware entity (a one chip secure microcontroller). It is capable of securely hosting applications and their confidential and cryptographic data (e.g. key management) in accordance with the rules and security requirements set forth by a set of well-identified trusted authorities [21].

The SIM has developed in a number of ways. For example, an NFC SIM embeds contactless NFC functionality into the secure environment of the SIM itself, allowing the phone to be safely used for payment at the point of sale.The SIM is used for other types of application such as mobile payment, in addition to the USIM application which provides network connectivity, thus enabling MNOs to enter the mobile payments ecosystem. As such, the SIM also has to pass the certification requirements of the Payment Schemes. This helps to produce interoperability and ensures that apps work across all platforms.

In addition to payments, the SIM also enables further applications, including industry moves to replace passwords with more secure alternatives. FIDO Alliance strong authentication specifications [22] support the use of embedded SIMs. GSMA's Mobile Connect initiative [23] relies on the SIM card for secure access and authentication for online services.

The SIM also protects application and service provider credentials and allows them to be stored independent of the network, for example in the bank usage of NFC. For all types of app providers, irrespective of industry, secure platforms today allow independent and securely stored co-existence.

Where no cellular connectivity is required in a secure environment and other forms of network connectivity are used instead, embedded secure elements (eSE) may be used to hold the application credentials and the network connectivity is secured by other means.

3.7 The role of the secure element management solution or service

The entity that operates the provisioning and management of secure services within the secure element is the secure element management solution/service or the Trusted Service Manager (TSM). It acts as the connection point between service providers, such as banks, transit operators and merchants, and the MNOs issuing the secure element [24]. Using both over the air (OTA) and over the internet (OTI) channels it accounts for life cycle management of services, applets, profiles and network access.

[20] http://www.gsma.com/aboutus/leadership/committees-and-groups/working-groups/fraud-security-group/security-accreditation-scheme

[21] http://www.gsma.com/aboutus/leadership/committees-and-groups/working-groups/fraud-security-group/security-accreditation-scheme

[22] https://fidoalliance.org/specifications/overview/

[23] http://www.gsma.com/personaldata/mobile-connect

[24] http://www.gsma.com/digitalcommerce/wp-content/uploads/2013/12/GSMA-TSM-White-Paper-FINAL-DEC-2013.pdf

3.8 Moving beyond today's security state of the art

Top security priorities in mobile communications today include protection against call fraud and data package interception, privacy, ensuring that users are who they say they are and that eavesdropping is not possible. Techniques employed include data encryption, key management and mutual authentication [25]. Much of the security enabled in 2G, 3G and 4G comes from the use of the secure, tamper-resistant nature of the SIM card, which enables security without providing usability disadvantages to the end user.

Priorities are likely to be broader in 5G, reflecting the far wider range of applications possible. Furthermore, these applications foreseen for 5G will place considerable technical demands over and above those present today that will create security challenges.

General security requirements from earlier generations are likely still to hold true for 5G but there will be differences nonetheless. In addition, topics considered but not adopted for previous generations like user data integrity, non-repudiation and IMSI catching may also be considered [26].

Significant factors in 5G will include higher speeds/lower latency combined with power efficiency needs, a wider variety of actors and device types, a greater range of threats and more use of the cloud and virtualisation. In order to avoid bottlenecks and integration difficulties, new security approaches will be needed. There will be a strong need for battery optimisation, particularly for IoT and M2M sensors.

At the same time, it may also be necessary to maintain backward compatibility with earlier generations.

Any solution must also take into account the needs of lawful interception, given that today law enforcement agencies are able to use the removable SIM to obtain information.

[25] http://www.ericsson.com/lb/res/docs/whitepapers/wp-5g-security.pdf
[26] Towards 5G Security, Horn & Schneider, Nokia, http://bit.ly/296sGDq

4. General introduction to security layers

4.1 Making sense of a layered model

Discussing security requirements in 5G is quite a complex undertaking. Security features may touch on and influence multiple security layers. Those security layers are independent of each other, but may be combined in order to realise the overall system security. Associating the requirements to the respective security layers helps to avoid confusion and to better derive potential solutions. In addition, a clear association of particular security mechanisms to the respective security layer may help to clarify the potential impact of certain regulatory requirements.

Across all building blocks / segments, security requirements will be based around: the consumer, devices, the network, the services and the backend. It is essential to strictly separate the security requirements and apply them to the relevant part to which they belong. We therefore propose in this document to split security into the following high level types of security layers:

- Network
- Service
- Application
- Device or user equipment (UE)
- Consumer.

This diagram shows how these layers can be further broken down.

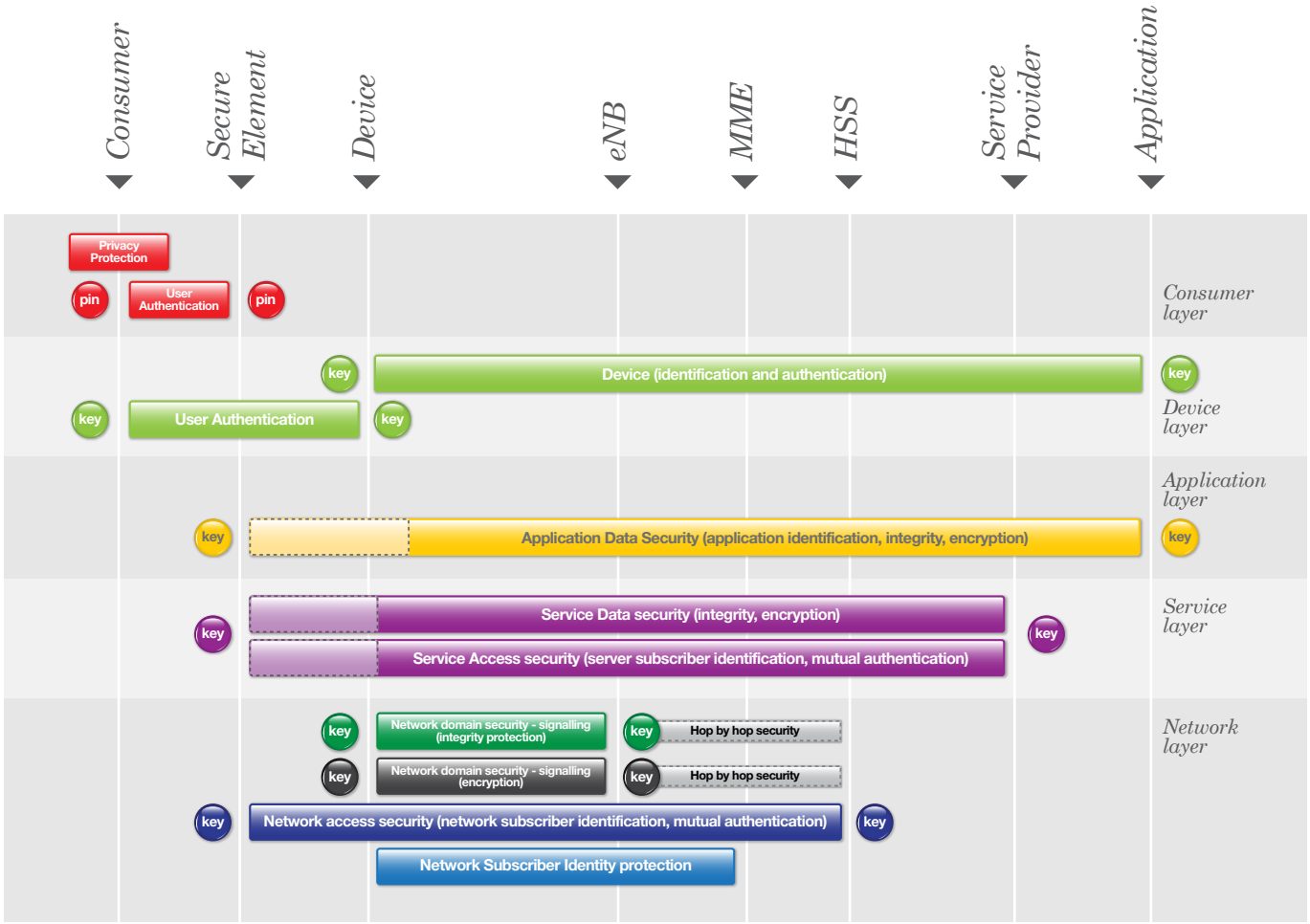

Fig 4.1 Security layers

4.2 Network layer security

This layer can be split into two parts: network access (part of the control plane) and network application (user plane).

Network access covers security features that make sure that only authorised and authenticated subscribers are allowed to gain access network functions. Different types of access, i.e. 3GPP or non-3GPP, as well as the ability to access different network slices are covered within this layer.

As subscription related data such as identification or location information are exchanged between the network entities when accessing the network, security mechanisms ensuring privacy are also considered within this layer.

The network application layer covers security features such as integrity protection and encryption of data that is transferred over the radio interface.

4.3 Service layer security

Services can be split into those that are defined by 3GPP, i.e. 3GPP services, and services that are provided by service providers / third parties.

As such, service layer security mechanisms are defined within the domain of the service provider and cover aspects such as service authentication, confidentiality, integrity protection and privacy. Those mechanisms may either be based on features provided by the network layer or may be completely independent of them.

4.4 Application layer security

Service providers implement their services by providing applications to their subscribers. In addition to the security provided by the service layer, each application may implement additional and/or different security mechanisms. These could cover security mechanisms such as end-to-end data encryption and integrity protection.

4.5 Device or UE security

Certain devices are required to implement security mechanisms in order to make sure only authorised users have access to device resources and in order to make sure that assets such as the device identifier cannot be manipulated. Those mechanisms are covered within the device security layer. In addition, aspects such as provisioning the UE with service or network access subscriptions, device theft, device integrity and grouping of devices (e.g. for bulk authentication and management) are covered.

4.6 Consumer security

The consumer security layer mainly covers privacy aspects. Privacy is a key requirement and has to be respected due to regulation and national privacy protection frameworks. This layer also covers consumer identification and authentication of the consumer towards other 5G system components such as the device.

5. Security requirements by segment - analysis and proposed solutions

5G brings a wide range of new functional requirements. It's important to understand how these requirements map across to security requirements. For example, a need to initiate communication quickly will affect the frequency and the way in which authentication and key agreement procedures can and therefore should be carried out[27]. This chapter therefore considers use cases and requirements, functional and security, by segment and recommends strategies and techniques for meeting those requirements.

5.1 Network operations

Underlying and horizontal to each 5G market segment is the operation of the 5G network. Existing usage patterns for mobile[28] in terms of data and downloads strongly suggest that operators will face a 5G future surge in data combined with a decrease in average revenue per connection (ARPC). As a result, they will look to cut costs. The solution increasingly appears to be Network Function Virtualisation (NFV) and network slicing, both for cost and technical reasons. In addition, NFV adds flexibility, scalability, decreasing power consumption, deployment of new and innovative functions with a shorter time to market and decreasing risks.

Indeed, 5G technology is expected to be built around a "network of networks" concept and real-time handover between the network technologies involved will be key to success. NGMN[29] identifies 50 Mbps throughput as an absolute minimum requirement (many envisioned applications require 1Gbps+) and the network will need to be optimised for data rate, latency, power efficiency and connection numbers.

Network operations also covers access to licensed or unlicensed, public or private networks. All these scenarios as well as the new features require security mechanisms that ensure authenticated access to the network both for regulatory reasons and for liability. They are applicable for all services utilising the 5G network.

5.1.1 Connection use cases

Use cases within network operations are envisioned by 3GPP to fall into the following families[30]:

- Flexibility
- Scalability
- Mobility support
- Efficient content delivery
- Self-backhauling
- Access
- Migration and interworking
- Security.

These use cases imply a set of technical characteristics, one of the most important of which is network slicing. Combined with NFV, it provides the scalability operators need in the face of growing amounts of machine traffic and an increasing number of vertical segments with potentially conflicting requirements.

Latency requirements are an important consideration. Without formal standards for 5G, there are no clearly defined requirements as yet. GSMA[31] identifies sub-1ms latency and >1 Gbps downlink speed as clear requirements for 5G that would form a clear step-change from what is possible with LTE-A. NGMN calls for 10ms for control plane latency and from 4ms to 0.5ms for user plane latency[32].

However, as latency decreases to ultra-low levels, tough decisions about the level of security will be triggered. While every use case merits high security, some security measures are not compatible with ultra-low latency at the sub-1ms level and adaptations may need to be made.

Mobile edge computing will also be significant – where ultra-low latency is needed, proximity is a key consideration. According to ETSI[33], this "offers application developers and content providers cloud-computing capabilities and an IT service environment at the edge of the mobile network. This environment is characterised by ultra-low latency and high bandwidth as well as real-time access to radio network information that can be leveraged by applications."

27 Towards 5G Security, Horn & Schneider, Nokia, http://bit.ly/296sGDq
28 http://www.statista.com/statistics/266488/forecast-of-mobile-app-downloads/
29 https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf
30 3GPP TR 22.864 V1.0.0 (2016-02) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers - Network Operation; Stage 1 (Release 14)
31 https://www.gsmaintelligence.com/research/?file=141208-5g.pdf&download
32 Requirements for NGMN KPIs and Requirements for 5G
33 http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

5.1.1.1 Network slicing concepts

Network slicing is a mechanism [34] that allows 5G to meet its diversity of requirements (e.g. flexibility, quality of service) by enabling operators to support multiple virtual networks, while benefiting from economies of scale that come from large scale physical network aspects. It allows operators to customise autonomous and independent networks to best meet the requirements of different market scenarios.

It is enabled by technologies such as NFV, Software Defined Networking (SDN), Cloud-RAN (based on centralisation and virtualisation of base station baseband processing), and mobile edge computing [35].

However, it also raises the possibility of a range of scenarios that any security mechanisms must take into account. These are listed by 3GPP SA3 as [36]:

- "Network function sharing.
- Access network sharing.
- Access from less trusted networks.
- Coexistence within a network slice with 3rd parties' network functions.
- Coexistence between network slices with different security assurance requirements.
- Simultaneous UE connections to multiple network slices.
- Simultaneous UE connections through different access technologies.
- Possible deployment scenarios and trust relationship between the network operator and the service provider, e.g. third party application server."

Slices may also communicate with each other i.e. inter-slice communication. This is equivalent to two independent, autonomous, isolated networks communicating with each other and is managed in the same way, unless functions are shared between the slices.

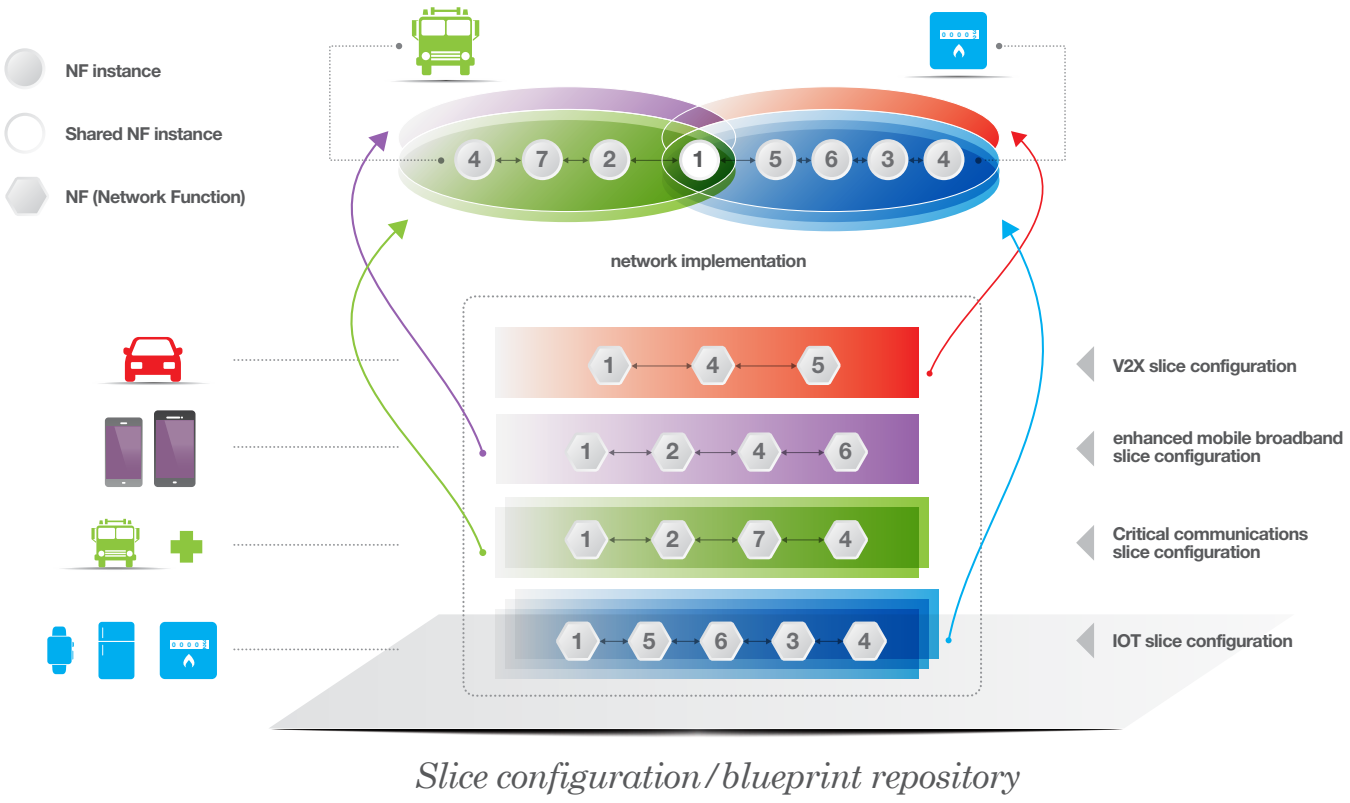

Figure 5.1 Networking slicing

[34] http://www.telecomtv.com/articles/5g/sk-telecom-and-ericsson-to-work-on-network-slicing-for-5g-12683/

[35] http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing

[36] 3GPP TSG SA WG3 Security Meeting #83, 9-13 May 2016, San Jose del Cabo, Mexico – S3-160798 – pCR: Key issues of security on network slicing

5.1.2 Security requirements

In turn, these use cases and resulting functional characteristics lead to security requirements which are examined in the following sections.

5.1.2.1 Network slicing security

The nature of slicing leads to a range of specific security requirements. 3GPP SA3 identifies these as[37]:

- "Security isolation of network slices.
- Security mechanism of each slice.
- Security on UEs' access to slices.
- Security on sensitive network elements.
- Security on management of slicing.
- Security on interacting with third party.
- Virtualisation security."

Network slices are intended to be independent and autonomous, which seems to imply security policies and configurations that differ according to functional needs of the slice. However rather than being a logical entity, a slice is a logical mapping of a set of functions. Some of those functions will be shared with another slice. Therefore, you cannot simply apply a security policy to a slice. Instead, what's really important is the access control, authorisation and authentication between individual virtualised functions (e.g. is this function allowed to talk to that one – i.e. are they on the same slice)?

As a result, each virtualised function requires its own authentication mechanism to be able to mutually authenticate other functions that it communicates with that are on the same slice (as defined in the forwarding graph of the slice).

In addition, the compromise of one slice should not be able to impact another slice. Nor should the compromise of a function within a slice affect any other slice. So any mission critical function, for example concerning subscription management or network authentication, should not be shared across slices.

Any subscription is protected by a NAA within the UE that takes care of network identification, authentication and encryption.

Several architecture models are under discussion and are possible. The authentication might be both at a network operator level (for a group of slices belonging to a same network operator) or at a service / application level i.e. at a slice level (as this service uses the network functions provided via a particular slice).

The UE might need to concurrently connect to multiple slices offering different services in parallel, meaning that multiple NAAs/ subscriptions have to be active concurrently.

5.1.2.2 Privacy

In the 5G ecosystem privacy and security are seen as complimentary[38], building on the inclusion of subscriber privacy from 2G onwards.

While privacy extends beyond technology into regulation, legal frameworks and commercial activities[39], there are security requirements that are important to observe. Many of these tie into the requirement explained above to use network slicing.

A temporary subscriber identifier (e.g. temporary IMSI) should be used to ensure pseudonymity between the 5G network and the UE.

In addition, ideally, the permanent identifier of each NAA should never be communicated between the network and the UE. This is because communications with the network could take place over a non-3GPP or a roaming network that is implicitly untrusted. Should the permanent identifier need to be transmitted, it should be sent protected using a suitable, secured mechanism.

The keys used to encipher voice/data in transit should be stored within each NAA and will be negotiated separately for each network slice to which the UE is connected.

To maintain forward compliance, there should be separate, independent key(s) for encryption and integrity protection. Keys for the different security layers outlined in chapter 4 should be independent of each other. This is important for data protection.

The network information (e.g. Network Measurement Report, neighbouring cells, …) for the subscriber/user for each slice should be accessible by the NAA. This information should not be accessible by the application processor of the device unless explicitly authorised by the entity operating the slice. A mechanism within each NAA could be made available to permit access to this information on a case by case basis (e.g. in the case of lawful interception).

[37] 3GPP TSG SA WG3 Security Meeting #83, 9-13 May 2016, San Jose del Cabo, Mexico – S3-160798 – pCR: Key issues of security on network slicing

[38] https://www.ericsson.com/res/docs/whitepapers/wp-5g-security.pdf

[39] https://www.ericsson.com/res/docs/whitepapers/wp-5g-security.pdf

5.1.2.3 Interworking

Interworking enables handover and mobility between different 5G networks, different network slices and 5G and earlier generation mobile networks. As such it creates a variety of security challenges.

It is possible, for example, for multiple NAAs and thus subscriptions to exist on the same UE.

Several different scenarios exist:

- Multiple NAAs and subscriptions (assuming each NAA holds one or more subscriptions) per a single slice.
- One NAA with one subscription per slice.
- One NAA for multiple slices.
- Multiple NAAs for multiple slices (combining the above options).

However, SIMalliance believes that sharing a single NAA across multiple slices is inadvisable, unless the NAA holds different subscriptions and associated configuration data. Mission critical elements should not be shared between slices as this produces a shared Home Subscriber Server (HSS) on the network side, leading to a compromise of one slice producing a compromise of others.

To accommodate fast handover between slices, a connection priority, starting from a default slice, should be specified to indicate to the UE which slices to connect to and in which sequence.

To protect against attacks such as denial of service, a mechanism should be made available for each NAA that can manage the access priority to a slice amongst the different UEs trying to connect to this specific slice. This could also be used for a maintenance engineer of a given network who is on call in case of network difficulties to gain priority access to the network ahead of regular subscribers.

Each NAA should contain a list of preferred 3GPP networks including an indication of the respective network slice as well as the connection priorities with regards to non-3GPP Radio Access Technology (RAT). These non-3GPP RATs could include WiFi and LiFi, both inside and outside the network operator's domain.

Subject to use case, the UE should systematically authenticate to all available network resources in case device mobility necessitates fast handover between multiple bearer network resources. This means multiple authentications at the same time to multiple slices or when migrating between slices. Speed of handover is also important for a consistent, seamless user experience.

This is likely to be more of a requirement in enhanced mobile broadband or critical communications, where devices will be highly mobile, than in massive IoT, where devices may well be stationary and so connecting to more than one bearer would simply be inefficient in energy and signalling terms.

Likewise, should a network operator wish to migrate a subscription from one network slice to another, this should be a seamless experience for both the consumer or IoT device where there is no service interruption and also for the network operator who simply migrates the data and credentials corresponding to the subscription from one slice to another.

Secure access to the remote provisioning system should be available at all times – please see the section on provisioning in 5.2.1 for the implications of this in massive IoT.

NAAs for 5G should enable connectivity to legacy networks that are not deprecated. In order to do this most securely, the authentication centre within the network slice should force the usage of the strongest authentication mechanism available to authenticate users to legacy networks.

5.1.2.4 Device identity protection

The device identity and the identity stored in the NAA should be separate and independent from each other, as is the case in earlier generations with the storage of the IMEI and IMSI/keys in separate logical entities. This supports mobility, security and flexibility.

Furthermore, the device's identity should be stored in a secured tamper-resistant entity to ensure that the identity cannot be modified as per today's IMEI modification loophole on some stolen devices.

5.1.2.5 Subscription protection

The NAAs should also be stored in a secured tamper-resistant entity in a way that ensures the separation called for above.

It should not be possible to clone or copy the credentials associated to each NAA within or from the UE, nor any part of the executable code protecting those credentials and the associated algorithms. Controls should be put on the network side to ensure that the same NAA is not connected multiple times to a local / roaming 3GPP or non-3GPP network unless explicitly intended by the mobile network operator.

5.1.2.6 Authentication, encryption and key management

Authentication should take place at the centralised / cloud RAN, wherever possible. In some scenarios within, for example, IoT, critical communications and enhanced mobile broadband, it may take place at the mobile edge. This is relatively straightforward for stationary devices. For mobile devices, the location must be known for the pre-calculated authentication vectors to be available locally at the mobile edge.

However, voice/data encryption could be distributed to mobile edge clouds in case of latency requirements where both parties communicating are connected through the same mobile edge cloud.

Keys corresponding to those provisioned in the NAA, as well as any appropriate network master keys or configuration parameters, should always be securely stored on the network side using a physical computing device that safeguards and manages digital keys. (i.e. a tamper-resistant entity within the network).

5.1.2.7 Geographical usage of a subscription

It should be possible to limit the geographical usage of a NAA to a given location or set of location identities.

A mechanism should be made available to capture the physical location of a UE at a given point in time and to securely store the location of the UE in a read-only manner within the NAA(s) corresponding to one or more network slices. The UE should then only allow connection to the geo-locked network slices where a fixed location is present in the NAA for that slice from that fixed location.

5.1.2.8 Security overheads

The device identification, NAA identification, authentication, integrity protection and encryption protocols should be optimised as much as possible to minimise the network attach and communication overheads without compromising the security.

5.1.2.9 Lawful interception

Depending on local regulatory requirements, there might be a need to store or access communications from mobile devices (either data or voice). This is complicated in 5G by the virtualised network core and the concept of network slicing, combined with mobile edge computing. It is therefore recommended that access to such communications requires strong authentication mechanisms and that any communications that are stored be confidentiality and integrity protected.

5.1.3 Security requirements allocated by security layer

The majority of security requirements listed within this section are of course related to the network security layer described in chapter 4. However, the building blocks on network operations also contain security requirements that are allocated to other security layers, as indicated in the table below:

	Network	Service	Application	UE	Consumer
Secure access to remote provisioning system	Access to the provisioning system may be provided by either using the 3GPP or a non-3GPP network	Provisioning is a service provided on top of the network access. Access to the provisioning server needs to be secured			
Device identity protection				Separation of device ID from subscription ID and secure storage of the device ID	
Secure storage of NAA				Requires the UE to provide a tamper-resistant entity	
Geographical usage and location information	Limit the usage to a geographical area			Securely store the location information within the UE	Protect the location information for privacy reasons

Figure 5.2 Network operations security requirements by layer

5.1.4 Complimentary recommendations

New types of interfaces between the device and the secure tamper-resistant entity, beyond that specified in the ISO specifications, may be needed, for example SPI or I2C, both of which are easier to implement for embedded device manufacturers and more power efficient than ISO.

The entity storing the NAAs, such as the secure tamper-resistant entity, must be capable of being (and should be) audited and certified by a third-party and functionally tested against a suitable industry-agreed functional compliance suite.

New form factors for the secure tamper-resistant entity may be needed to address the diverse needs around the various 5G sub-segments. As part of this the future may see a further evolution of the system on chip approach, i.e. iUICCs, which may lead to more deeply integrated and smaller form factors.

Firmware upgradeability should be a mandatory feature both on the entity hosting the NAAs. It should also be mandatory on the device part which is interacting with the NAA.

SIMalliance also proposes the use of longer symmetric keys in preference to asymmetric approaches, particularly in view of the projected lengthy lifespan of many IoT devices. This aligns with the recommendations made in ETSI's Quantum Safe Cryptography and Security white paper [40].

5.2 Massive IoT

The Massive IoT segment is extremely broad, covering not just M2M but consumer IoT too. It is likely to consist of an extremely diverse ecosystem of potentially very simple devices such as sensors or trackers, existing machine-to-machine type devices such as power or water meters alongside advanced consumer devices such as smartwatches and other wearables.

There is a degree of disagreement about how quickly this sector will turn to 5G. Some projections suggest that there may be anything from 20.8 billion [41] to 50 billion [42] connected devices by 2020. While not all of these will use 5G connections, Gartner's suggested [43] CAGR of 30% for 5G implies that at least 40% will be 5G compatible.

Clearly these figures stand in contrast to the much lower Ovum figures quoted in the introduction, but Ovum also suggests that initial 5G connections are likely to come from the enhanced mobile broadband segment and that IoT is not likely to utilise 5G before 2021.

The extreme broadness of this segment and the large variety of potential uses cases imply the following potential characteristics of the devices associated with the sector:

- **Administrable:** While some IoT devices will be very simple, others will be multi use case, complex or expensive. It may also be necessary to upgrade the device and applications embedded in them post issuance.
- **Configurable:** Remote configuration may also be required. This is likely to involve smaller amounts of data than remote administration.
- **Connectivity:** Connectivity requirements may differ by use case. Some IoT devices will require a permanent connection (e.g. for home security monitoring), others only an occasional connection, used only when data needs to be uploaded to a server (e.g. an "Alert" use case). The connection can be direct on a 3GPP-RAT or through a gateway on a non-3GPP-RAT. In some sectors, primarily industrial, networks will be walled gardens, with external connectivity enabled only to one server in the service provider's cloud.
- **Position:** Some IoT devices will remain in the same location and never move. In this case the device will be connected to the same network. Others will move and so the ability to manage network handover securely will be required.
- **Network:** To assure the best connection and usage, the IoT device may be:
  - Latched to the global network (Public 3GPP-RAT).
  - Latched to a dedicated 3GPP-RAT (Private 3GPP-RAT).
  - Connected through a non-3GPP-RAT.
  - Managed over an ad-hoc connection (Device2Device).
- **Group:** In some cases, IoT devices will be associated to one or more groups of devices. This enables efficient management of those devices and allows:
  - Identification of a group of devices.
  - Update of a group of devices.
  - Authentication of a group of devices.

40 https://portal.etsi.org/Portals/0/TBpages/QSC/Docs/Quantum_Safe_Whitepaper_1_0_0.pdf
41 http://www.gartner.com/newsroom/id/3165317
42 http://iq.intel.com/how-5g-will-power-the-future-internet-of-things/
43 http://www.gartner.com/newsroom/id/3165317

5.2.1 Provisioning

In order for a device to connect securely to the network, it must possess NAA credentials. The process of introducing these credentials into the device is known as provisioning.

This raises the question of how you authenticate the device to confirm that it is a valid 3GPP device and eligible to download a 3GPP subscription. This is particularly problematic with devices that do not have user interfaces, where it may not be easy to tell if the device is online or reachable by the network. SIMalliance considers this a priority to define.

5.2.1.1 Bulk provisioning during manufacture

Bulk provisioning or pre-provisioning takes place during device manufacture. This type of provisioning allows deployment of the IoT device with credentials already embedded and ready to connect out of the box.

The IoT device must have a secure mechanism embedded for the loading of credentials in manufacture during the personalisation process. This secure mechanism can be adapted according to the security level of the manufacturing process. One option to realise this is to embed already personalised eUICCs during the device manufacturing process.

5.2.1.2 First use provisioning

This type of provisioning permits the sale of the IoT device without network specific credentials. Therefore, there is a need to be able to provision credentials after sale but before first use.

In this case the end-user or an authorised entity needs the ability to place the initial credentials into the device through a secure process, using either dedicated hardware or through an internet connection and an online process (for example as set out in the GSMA's remote provisioning architecture [44]). The management is defined by some basic operations such as:

- Load a new NAA (AKA credential).
- Delete a NAA.
- Activate (select) a NAA.

However, 3GPP SA1 has proposed a requirement [45] to be able to connect out of the box with zero configuration without the need for a pre-installed 3GPP subscription. So the means of making this initial connection must be assured and the technology used to protect the download of the NAA must be defined.

For example, there is the question of where the device will initially connect. How will it be re-routed by whatever server it connects to into the correct NAA management entity for the contract the OEM/SP has secured? This requires business logic specifying a triangulation of the type of the device, the service provider and the device location.

This suggests an approach similar to that taken in the GSMA architecture. However, it should be noted that the GSMA architecture may need to be adapted to fulfil the requirements defined for 5G.

5.2.1.3 Other credentials

The device must also provide secure storage for credentials required at the application layer, providing security to the service provider as well as the MNO. For example, some IoT devices may need to store payment credentials. Provisioning must therefore take into account these too.

5.2.2 Connection use cases

This section describes different ways in which an IoT device can connect to 3GPP networks. The first is a classical schema that will be common across all segments. However, it is also possible that relay devices may find a more limited use in critical communications and enhanced mobile broadband too.

[44] http://www.gsma.com/newsroom/all-documents/sgp-02-v3-1-remote-provisioning-architecture-for-embedded-uicc-technical-specification/

[45] The potential requirement is in 3GPP TR 22.861 (3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers for Massive Internet of Things) available here: http://www.3gpp.org/ftp/Specs/archive/22_series/22.861/

5.2.2.1 Direct 3GPP-Radio Access Technology connections

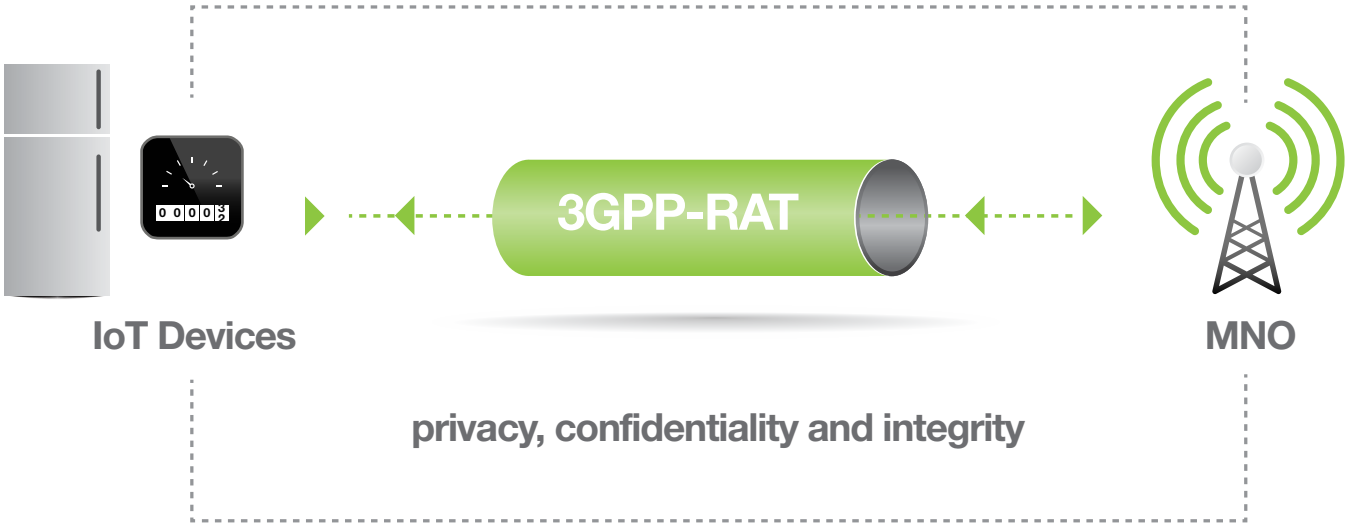

Fig 5.3 A direct 3GPP-RAT connection

In this schema, the device connects directly via a 3GPP-Radio Access Technology (RAT) network. A typical example might be a smart meter. Key objectives are to protect the privacy, confidentiality and integrity of the data transmitted.

5.2.2.2 Using a relay device through a 3GPP-RAT connection

In this schema, the IoT device contains 3GPP credentials but is not connected to the 3GPP-RAT network directly. Instead it connects to network core via another independent non-3GPP RAT network to a relaying device which then connects via a 3GPP-RAT. In effect the relay UE acts as a switch. A typical example might be a smart watch connecting via Bluetooth to a mobile phone acting as a relay. However, what is important is not the type of connectivity but the trust between devices. This configuration allows enhanced overall security of the connectivity by not relying on the local security provided by the non 3GPP RAT between the device and the relay UE.

Clearly one reason for this type of use case is to manage device cost. Through providing 3GPP connectivity to devices that might not otherwise have it, it also allows MNOs as well as just the OEM to provide services to such devices. It also enables lawful interception.

Two cases may need to be distinguished:

- The IoT device has its own credentials.
- The IoT device shares the credentials of the relaying device.

Again, it's important to protect the privacy, confidentiality and integrity of the data transmitted.

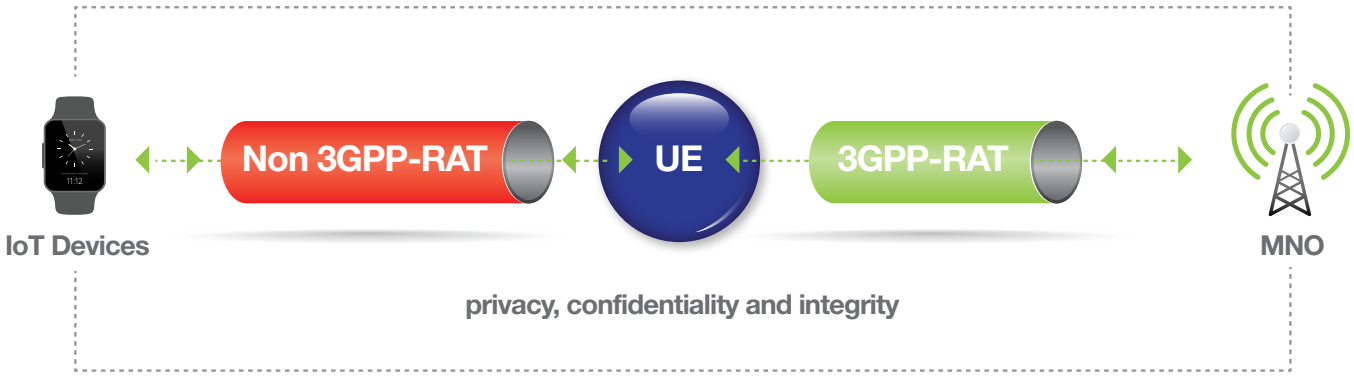

Fig 5.4 A non-3GPP-RAT connection to a relay and then via a 3GPP-RAT

5.2.2.3 Using a relay device through a non- 3GPP-RAT connection

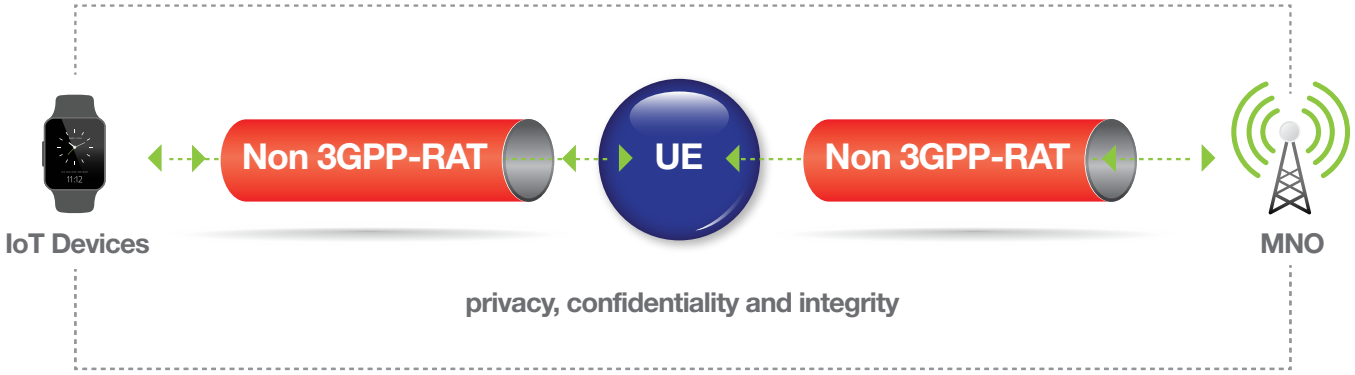

Fig 5.5 A non-3GPP-RAT connection via a relay

The last statement is equally important in the third variant which will see the relay itself connect to the backend via a non-3GPP-RAT. While it is debateable whether this is in scope for 3GPP standardisation and hence this document, it nonetheless involves the use of 3GPP credentials. A typical example might once again be a smart watch connecting to a mobile phone acting as a relay but in this case the phone itself connects to the network via WiFi.

5.2.3 Security requirements

The broad range of use cases in this sector open up the possibility of a broad range of threats and attack vectors. The simpler the device, the greater the risk that it will attract attacks. A simple device does not necessarily require simple security; a high level of security will be necessary across the whole massive IoT sector.

5.2.3.1 Credentials repositories and algorithms

Irrespective of the connection use case, to record all types of credentials (3GPP-RAT or non-3GPP-RAT), the IoT device should own a secure and tamper-resistant repository.

To manage these credentials, the IoT device must support a secure means, such as the Subscription Manager – Secure Routing (SM-SR) [46] to Load, Delete and Activate, from the repository in which profiles are stored, to the device into which they are being downloaded.

At the same time the NAA within the IoT device must be capable of storing and executing algorithms requested for the AKA authentication of the 3GPP-RAT.

For non-3GPP-RAT, the IoT device must support network requirements. Any non-3GPP credentials should be stored using the same level of security as for the 3GPP NAA credentials themselves. The device should also be able to support the authentication algorithm according to the network type (WiFi, Bluetooth, LiFi, or others).

5.2.3.2 Subscription renewal

There are times when it may be necessary to change NAAs or credentials on the IoT device during its lifecycle, for example for resale, or for security reasons such as hacking. To trigger this renewal operation, an external event must take place. The event can be the reception of new credentials (pushed from a server) or just an order to regenerate internally a new credential. This is most likely to affect 3GPP credentials.

[46] http://www.gsma.com/connectedliving/embedded-sim/how-it-works/

5.2.3.3 Power consumption

Low power consumption may be a critical requirement in some areas of massive IoT, irrespective of the connection type.

To implement this requirement, a new low level standard interface allowing a state of 'zero' current consumption i.e. a hibernate state must be defined. This has security implications.

Either, at each point the IoT device needs to go into a hibernate state, the connection session data must be recorded in a persistent secure area before going into the hibernate state.

Alternatively, a new connection can be restarted (with re-authentication and new session data generated) when the IoT device leaves the hibernate state. Given that some IoT devices go online relatively rarely compared to devices in other segments, a policy could also be created to align and make proportional authentication frequency to frequency of data sending and quantity and sensitivity level in order to manage power consumption.

The more efficient generation of session keys and more efficient algorithms will also reduce energy consumption.

5.2.3.4 Secure area for end-to-end (application) protection

As the main function of an IoT device is to transfer application data between the device and the server, a secure service must be embedded as a separate data protection function, independent of any relationship with 3GPP or non-3GPP-RATs. This will take place at the application layer (please see figure 4.1.)

This service will secure data according to the level of security required (cipher and/or signature) and will be based on symmetric or asymmetric keys according to the IoT device use case and network bandwidth. Given the arguments advanced by ETSI in its Quantum Safe white paper[47] about how symmetric algorithms using longer key lengths will better stand the test of time, SIMalliance recommends this as an approach.

5.2.3.5 IoT device identity

To identify an IoT device as unique in the ecosystem, it is necessary to define a unique identification number. This identification number may be the aggregation of several pieces of sub-information that:

- Identify the manufacturer of the device (static).
- Act as an identifier, like a serial number, during production (static).
- Identify a group of devices (modifiable).
- Act as a unique identifier for usage (modifiable).

The device identifier is security relevant and must be protected by secure storage against unauthorised modification.

5.2.3.6 Device location

There may be a business requirement for some IoT segments to restrict the usage of a given device to a certain location. Such use cases requiring such a restriction could be the smart home or a private virtual network within an industrial environment. It is recommended to store securely the restricted location list / identities (e.g. GPS coordinates or cellular network location information) and that only entities with relevant privileges are able to modify the location.

The location information stored must come from a trust-worthy source and could possibly be cross-checked at network level (e.g. cross-checking cellular network location information with GPS coordinates).

5.2.3.7 Integrity protection

In traditional cellular networks the biggest concern has been eavesdropping of communications (data or voice). With the arrival of massive IoT use cases, concern is likely to switch to integrity attacks whereby a user reverse-engineers an instruction carried out on an IoT device to perform an operation (e.g. open a door). Another type of concern is a denial of service attack (e.g. restricting a user from opening their door). Therefore, integrity protection is required.

5.2.3.8 Service layer security

At the service layer, service providers need to protect assets and offer trusted services, while managing the overall cost of development and deployment. Key security functions they will require include authentication functions (signing and verifying signatures), confidentiality functions such as encryption/decryption and key management.

[47] https://portal.etsi.org/Portals/0/TBpages/QSC/Docs/Quantum_Safe_Whitepaper_1_0_0.pdf

5.2.4 Security requirements allocated by security layer

Security requirements within IoT can be allocated at all levels of the layer model, as below.

	Network	Service	Application	UE	Consumer
Pre-provisioning during manufacture	No network access needed		Local provisioning during production could be seen as an application that is end-to-end protected	UE needs to be capable of being personalised during production (off-line personalis-ation). Tamper- resistant storage of credentials	
First use provisioning	Authenticated network access needed	Authenticated access to provisioning service needed	End-to-end security for provisioning application needed	Tamper- resistant storage of credentials	
Relaying the IoT device	Requires identification and authentication of both the relayed device and the relaying device	Relayed device may contain service NAAs and use them to authenticate to a service (e.g. fitness tracker service)	Relayed device may apply end-to-end security	Tamper- resistant storage of credentials	
Credential storage				Tamper-resistant storage of credentials. Secure execution of authentication algorithms	User authentication
Subscription renewal	Authenticate network access	Authenticated access. Trigger the renewal process	End-to-end security	Tamper resistant storage and switching of credentials	Correct association of subscription to consumer
Power efficiency	Adequate authentication frequency need to be applied. Efficient re-authentication needed. Efficient generation of session keys			*UE needs to go into hibernate state* (48)	*Seamless user experience*

Figure 5.6 Massive IoT security requirements by layer

(48) Italics indicates functional requirement with security implications

5.2.5 Complimentary recommendations

As discussed at the beginning of this section, massive IoT presents a broad range of attack surfaces, with devices at the simpler end of the spectrum potentially far more vulnerable.

As a result, far from allocating simple security mechanisms to simpler devices, it is these in particular that will benefit from the presence of a physical, secure, tamper-resistant environment. Because many of these simple devices will have both a projected lifespan of as much as 15 years and only periodic connection to the network and hence oversight and upgrade, it is vital that their security is built to last. Rather than considering the standalone purchase cost of the device as a metric for its security, it is far preferable to consider its lifetime value within the system, particularly given that longer term costs of making a short term decision on security cannot be known initially. Lower security today may incur hidden costs tomorrow.

Given the nature of these devices, it therefore makes sense to store AKA credentials in a physical, secure, tamper-resistant entity. According to ABI Research [49], "This is why hardware-based embedded security solutions play an important role in addressing the needs of these resource-constrained connected devices."

Furthermore, any technology recommended for use in this sector should have a 15-year lifespan. SIMalliance also proposes the use of longer symmetric keys in preference to asymmetric approaches, particularly in view of the projected lengthy lifespan of many IoT devices. This aligns with the recommendations made in ETSI's Quantum Safe Cryptography and Security white paper.

At the service layer, tamper-resistant hardware will help service providers meet their security requirements. Defining a standardised API for service layer security will ease and abstract the usage for service providers, removing the need for knowledge about its communication protocol or behaviour. Hardware providers can offer a large set of libraries focusing on high level functions abstracting the complexity of the tamper-resistant hardware for storage or confidential exchanges of data with the cloud. oneM2M, a global partnership project defining service layer specifications for M2M, has started a work item on Secure Environment Abstraction [50] for this reason.

New efficient algorithms, authentication policies and protocols that take into account lower power consumption should be evaluated for the IoT space for devices extremely limited in resources such as low-power sensors transmitting machine data at infrequent intervals.

5.3 Critical communications

5G networks are likely to play an even more fundamental role in critical infrastructure than previous generations did. They will participate in what will be a highly complex ecosystem, involving drones and their control, cloud driven virtual reality, smart multi-node factories, cloud driven robots, public safety, transportation and e-health. In many critical communications use cases human lives will be at stake, making functional and security requirements even more vital.

While these different use cases will present different security needs, the critical communications sector overall will be characterised by a high level of security and will have a wide range of overlapping functional requirements regarding:

- **Reliability** – This covers the certainty that data is received and decoded correctly within a given timeframe [51].
- **Availability** – This is closely related to reliability and covers equipment uptime, network capacity and coverage.
- **Throughput** – Refers to the amount of data that can be moved through the network in a given amount of time.
- **Ultra-low latency** –The amount of delay on the network between input and output - some areas of critical communications will require sub 1 ms latency [52].
- **Security performance** – Appropriate security for the use case in question, given differing requirements around latency, data quantity and frequency and other factors. This may mean that activities such as identification, authentication, encryption and integrity checks take place either on the mobile edge or at the core of the Cloud-RAN.
- **Data integrity and confidentiality** – Where human lives are concerned, protecting data and preventing any type of manipulation is vital.

Connections in general will be direct to the 3GPP network but there is possibly a use for relay devices in emergency networks to provide emergency access to the network during disasters. Please see section 5.2 for how this works.

[49] https://www.abiresearch.com/press/pressure-mitigate-iot-device-cyber-threats-mounts-/
[50] http://onem2m.org/technical/latest-drafts (see TS-0016 Secure Environment Abstraction)
[51] http://networks.nokia.com/sites/default/files/document/5g_requirements_white_paper.pdf
[52] https://www.gsmaintelligence.com/research/?file=141208-5g.pdf&download

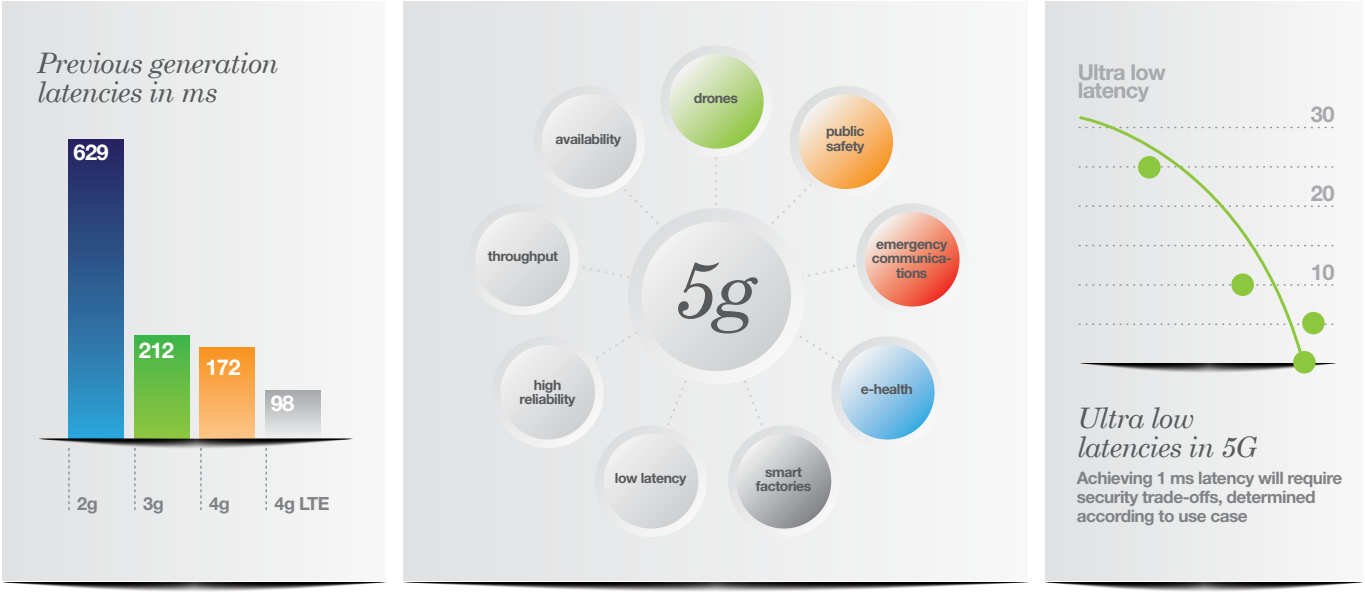

Figure 5.7 The critical communications segment

5.3.1 Provisioning

As explained in 5.2.1, a device cannot connect securely to the network without first obtaining credentials. As in massive IoT, some devices may be pre-provisioned, whereas others must be capable of securely downloading network access credentials before first use.

5.3.2 Mobile edge computing

The mobile edge computing aspect is key for critical communications to increase responsiveness of the network, with highly time critical functions being processed at the mobile edge. Less time critical items that do not need to be processed in near realtime are delegated to the core.

Close coordination between the edge and the cloud is therefore necessary, including in the security aspects where some authentication mechanisms could then be delegated to the cloud for instance.

For example, key updates must take place in parallel in the cloud and the edge.

5.3.3 Connection use cases

This section describes the different types of use case for critical communications.

These use cases imply several functional security requirements and hence different types of communication, in particular:

- Optimised security for reliability and best throughput of exchanged information.
- Fast security for ultra-low latency communications.

According to 3GPP [53], use cases that require high reliability may involve closed loop networks of limited size, such as factory automation in power plants. Here, a large and densely clustered number of sensors may communicate regularly over a short range with a controller. Process automation on the other hand may involve a similarly dense network of sensors but using open-loop communications.

Other examples may involve use cases that involve human operators, for example drones. These devices must be controlled quickly and reliably, but the involvement of human reaction times limits the latency that is required, for there is no point in striving for ultra-low latency where the human factor sets a ceiling for what needs to be achieved.

[53] 3GPP TR 22.862 V1.0.0 (2016-02) 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; Feasibility Study on New Services and Markets Technology Enablers – Critical Communications; Stage 1 (Release 14)

Ultra-high reliability may be a required feature in mission critical communications, which may also require preferential handling. These include emergency services communications where in some cases human lives are at stake, mobile healthcare, real time vehicle control, national security communications and industrial control scenarios.

Very low latency is a requirement for applications such as tactile internet, where a human machine or robot operator receives both visual and tactile feedback from a device, for example in virtual reality or in remote healthcare monitoring and treatment. Another description for this is extreme real-time communications. Clearly, the need for low latency does not negate the requirement for high reliability too.

Because of the speed and latency requirements of this segment, it's important to consider at what point between the C-RAN and the mobile edge identification, authentication, encryption and integrity checks take place.

In addition, a high speed algorithm may need to be recommended. Re-authentication policies should also be defined (e.g. after n events – 10 calls / 100mb etc), including how they work while a critical event is taking place.

5.3.4 Security requirements

For all use cases in this segment, the network authentication policy should evolve to meet critical communications requirements, in particular for the users to be able to transmit data anytime with a very low latency. In some use cases, for example public safety / emergency situations, the latency requirements may be such that authentication requirements may be reduced however confidentiality and integrity protection should not be compromised.

For the smart factory (private virtual network) use case, the management of the identification, authentication could be dynamically managed by the virtual operator of the network within a dedicated network slice environment, hosted locally at the edge or in a central cloud.

Some applicative security mechanisms could be envisaged in addition to the mechanisms provided by the network technology enabler.

5.3.4.1 Optimised security for reliability and best throughput

It is necessary to create confidentiality for device identifiers such as the International Mobile Station Equipment Identity (IMEI), or its 5G equivalent. These need to be integrity protected when being stored and transmitted, or anonymised if confidentiality cannot be set up.

Subscriber identity confidentiality also needs to be suitably protected. Moreover, in decentralised trust models, where devices cooperate to pass data between two endpoints through a series of intermediate hops, consideration for improved pseudo identifiers needs to be given in order to provide for non-linkability of subscriber and device identities.

Additionally, trust in all cooperating nodes will be required. This trust can come from the tamper-resistant secure element endorsed, qualified and controlled by the connectivity provider (i.e. operator). In critical communications, public lives are at risk and hence security of the NAA cannot be compromised.

To combat threats of passive attacks, mainly in the case of subscriber key compromise, provision of forward secrecy is to be considered for enhanced communication confidentiality, in the sense that an attacker who records some communications and discovers the long term subscriber key later, cannot go back and decipher the recorded communications.

5.3.4.2 Fast security for ultra-low latency

Low end to end latency, including the latency of initiating communications, will be an important shaping factor for security design, with any solutions having to focus on speed as well as effectiveness.

Some use cases, for example in temporary public emergency networks, require ultra-low latency. This requirement could impact the overall security in particular related to authentication procedures and end-to-end or hop-by-hop encryption.

Re-authentication happens in today's networks after a specified timeout period or a certain number of events such as calls or SMSs or data traffic.

Going forward the re-authentication policy should evolve to not delay any communications on the user plane of the network during critical communications. Re-authentication must never get in the way where human lives are at stake, unless aimed at rogue devices on the network. Therefore, a similar set of rules need to be defined.

Another central challenge faced by critical communication systems is the maximal rates of transmission of secured messages as well as their maximal key-generation rates of secured keys and rates of transmission of secured keys.

5.3.5 Security requirements allocated by security layer

	Network	Service	Application	UE	Consumer
Protection of device identifiers	Integrity protection at rest and in transit; privacy protection			Tamper-resistant storage of device ID	Privacy protection
Separation of device ID and subscriber ID	Identifiers' pseudonyms are used in transit			Tamper-resistant storage of device ID	Privacy protection
Resist passive attacks	Forward secrecy			Tamper- resistant storage of credentials	
Low latency	Optimised re-authentication; fast handling of security procedures;	Additional service layer security may slow down processing	Additional application layer security may slow down processing	Dedicated crypto-HW to support fast processing of algorithms	

Figure 5.8 Critical communications security requirements by security layer

The table above categorises the requirements listed previously and allocates them to the respective security layers.

It has to be noted, that dedicated use cases such as a closed loop factory are not covered in this table. Those use cases put requirements such as end-to-end protection and dedicated service layer authentication (e.g. using a factory owner's ID system) in particular on the service and application security layers.

5.3.6 Complimentary recommendations

By its very nature, with human lives at stake in some use cases, critical communications as a segment requires the highest level of security protection. Both the subscriber identity and the device identifier therefore need to be protected against malware or unauthorised applications when being transmitted over the network and when being stored. As a result, best-in-class security is needed, meaning tamper-resistant hardware, potentially with enhanced resources, to meet the ultra-low latency requirements.

In order to manage identities on the network, privacy management, group management and user and device authentication are also necessary. There may be several layers of encapsulated authentication required at both network access and service levels and authentication may be required much more frequently than in segments such as IoT.

Some use cases, for example in temporary public emergency networks, require ultra-low latency. This requirement could impact the overall security in particular related to authentication procedures and end-to-end or hop-by-hop encryption. However, in most critical communications situations, for example in remote surgery, the highest level of security is necessary and hence latency can be optimised using a variety of techniques.

It's also important to decouple the authentication on the control plane from the encryption on the user plane.

That implies that there should be no critical path present on the user plane. Since binding the derivation of keys for user plane encryption to the control plane could cause a service lapse, if a re-authentication were to be triggered during a critical communications session, such binding must also be avoided too. Furthermore, the authentication mechanism on the control plane must not impact the latency for critical communications taking place on the user plane.

In addition, instead of starting to encrypt traffic when the encryption key has been established, the transmission can potentially be started with integrity protection included in the first message. Suitable coding techniques will ensure that only legitimate receivers are able to process an integrity protected message, with the connection dropping if integrity protection fails.

For faster handling of security procedures consideration should be given to ideas such as establishing shared keys between entities in anticipation that they may need to communicate. Such communication may also take place over non-3GPP networks or through a mobile cloud without traffic going through the central cloud RAN.

5.4 Enhanced mobile broadband

This 5G segment encompasses the use of tablets, mobile phones and other portable devices. In fact, it offers a range of very diverse scenarios, with indoor and outdoor use, fast moving versus slow moving devices, use on trains and planes, hotspots such as offices, crowds and high density areas, low density areas and varying levels of network capacity requirement, depending on time and other factors. As such, it presents very similar security requirements to the underlying network operations segment.

5.4.1 Use cases

Enhanced mobile broadband covers use cases which fall into the following sub-segments:

- Higher data rates
- Higher density
- Higher user mobility
- Devices with highly variable data rates
- Broadcast services
- Different deployment and coverage scenarios.

These sub-segments are extremely varied with the end users indoors or outdoors, above or below ground, and in urban or rural areas. However, there are specific features related to mobility and volume of data, which highlight important differences between each sub-segment.

5.4.1.1 Higher data rates

The main focus of higher data rates use cases are data rate requirements for peak user experience and downlink/uplink, but always with UEs relative speed to ground up to 10 km/h.

Typical scenarios might involve users making a real-time video meeting in an office or in the street, frequently uploading and downloading data of varying quantities from servers, multi-media traffic towards the internet or device to device communications, broadcast transmission such as 4k UHD or 8k UHD video streams and running 8k 3D video streaming for uplink and downlink. In these cases, efficiency and reliability are important in order to be productive; also this scenario should support residential deployment with a latency of [10 ms].

5.4.1.2 Higher density

The requirements will be different for scenarios with higher density, such as high traffic density, when there is a high volume of data traffic per area, and high connection density, when the data is transferred for a high number of connections. End users are expected to be in a densely populated area, but always with UEs speed relative to ground of up to 60 km/h. Primary use cases are: handling high resolution real-time video conferences, uploading and downloading high volume and high capacity multi-media traffic towards internet in stadiums or in dense city centres, frequently uploading and downloading a very high volume of data from servers and using interactive applications.

5.4.1.3 Higher user mobility

There are three general scenarios under the umbrella of higher user mobility: seamless enhanced mobile broadband in fast moving vehicles (up to 200 km/h), in fast moving trains (up to 500 km/h) and enhanced connectivity services in fast moving airplanes (up to 1000 km/h).

Typical scenarios enable enhanced navigation for users through instant and real-time information, and to get access to high quality mobile internet connections for entertainment, work, interacting with social clouds, or infotainment. Furthermore, this group covers enhanced navigation through instant and real-time information for safety and vehicle diagnostics.

5.4.1.4 Variable data rates

Devices with highly variable data rates includes use cases where smartphones have multiple applications which frequently exchange small amounts of data with the server side of the application but where larger amounts of data are needed only occasionally. Typical examples are smartphones with applications that often exchange location updates and notifications, but rarely download / streaming or smartphones serving as the gateway to wearable sensors, sending small messages on a periodic basis. In these cases, one of the main approaches for the network is to be flexible, in order to provide efficient service no matter the amount of data in traffic and to avoid any negative impact to battery life for the device and minimise use of signalling resources.

5.4.1.5 Deployment and coverage scenarios

In deployment and coverage scenarios, UEs must have speed relative to ground of up to 120 km/h. Small area connectivity covers scenarios where the users and their serving nodes are expected to be deployed indoors, such as real-time video meeting and frequently uploading and downloading data of different sizes from the company's servers. In that case, wide area connectivity scenarios are based on providing seamless mobile broadband services to users, such as mobile cloud office, online games/ videos, and augmented reality. For wide area connectivity satellite based access could complement terrestrial based networks.

5.4.1.6 Broadcast services

It is envisaged that using the high speed connectivity that 5G brings, live broadcast television, special events or video on demand services could leverage the 5G network. Security to control access to such content to satisfy broadcast permissions / rights restrictions should be implemented in this space. Existing standardised key management systems for broadcast services (e.g. Multimedia Broadcast Multicast Service [MBMS], Open Mobile Alliance Mobile Broadcast [OMA-BCAST]) should be reused for this purpose in 5G.

5.4.2 Security requirements

Unlike massive IoT and critical communications, enhanced mobile broadband brings virtually no additional segment specific requirements. These requirements are likely to mirror those in the network operations segment, with almost no modifications.

However, there is one major difference in the use cases related to highly variable data rates, where the MSISDN, the number that identifies the subscription to the network, needs to be able to be transferred from one device to another in this segment (e.g. from a smartphone to a smartwatch) unlike in the other segments, where devices are just machines requiring data connectivity that do not care about their identifiers. This process should be a secure one so that one user cannot take over the MSISDN of another user without their consent.

5.4.3 Security requirements allocated by security layer

These requirements are likely to mirror those in the network operations segment, with no modifications.

5.4.4 Complimentary recommendations

This solution is likely to mirror that in the network operations segment, with no modifications except that in this segment there must be a secure process, which could be proprietary to the mobile network operator, to prevent one user from taking over the MSISDN of another user without their consent.

5.5 V2X

Over recent years, vehicles have become more and more 'connected', through in-car diagnostics, maintenance and entertainment. For example, today approximately 90% of BMWs count as connected cars[54].

Moving forward, autonomous vehicle efforts from Google and Tesla are driving this trend too, as is interest from transit operators, for example the trial in Helsinki of self-driving buses[55].

The objective is to make driving safer, more comfortable and more efficient, especially in ecological terms, according to NGMN[56]. It also predicts that "highly automated driving will hit the road around 2020, and mature towards 2025-2030."

This has led to new communication requirements for vehicles as autonomous vehicles must be consistently aware of, and able to interact with, what surrounds them.

For example:

- V2I (Vehicle-to-Infrastructure) – obtaining data from road signage and signals[57].
- V2V (Vehicle-to-Vehicle) – to avoid collisions with other vehicles.
- V2P (Vehicle-to-Pedestrian) – ensuring safety with pedestrians and cyclists.
- V2N (Vehicle-to-Network) – real time data about traffic.

In this segment, there will be functional and security requirements in common with other 5G segments including enhanced mobile broadband and critical communications. Furthermore, many of the objects autonomous vehicles interact with, such as road signage, parking meters or road charging infrastructure, will form part of the internet of things.

Already, connected vehicles provide attractive targets to attackers, even if reported hacks, such as the Chrysler Jeep hack in 2015[58] where the hackers were able to force a Jeep off-road and into a ditch, are limited in scope and have been carried out by researchers or academics.

In addition, like critical communications, V2X encompasses scenarios where human lives are at risk.

54 http://europe.autonews.com/article/20160614/ANE/160619974/bmw-exec-says-industry-ready-to-battle-hackers-and-make-move-to-5g

55 https://www.theguardian.com/technology/2016/aug/18/self-driving-buses-helsinki

56 https://www.ngmn.org/uploads/media/160610_NGMN_Perspectives_on_Vertical_Industries_and_Implications_for_5G_v1_0.pdf

57 gsacom.com/paper/cellular-vehicle-everything-qualcomm-presentation/

58 https://www.theguardian.com/technology/2015/jul/21/jeep-owners-urged-update-car-software-hackers-remote-control

5.5.1 Use cases

5.5.1.1 Autonomous/ cooperative driving

The car needs to understand and interact with its environment. That includes detection of pedestrians, interaction with other vehicles to anticipate collisions and taking smart decisions to avoid traffic, make emergency stops or adapt cruise control in order to improve road congestion and safety of road travel.

These requirements apply whether during the sort of driving conditions we would recognise today or high density platooning, where large numbers of autonomous vehicles may drive very close to each other and all vehicles must be able to respond in synchronicity.

5.5.1.2 Tele-operated driving

In this use case, vehicles could be viewed as drones on wheels. Automated Emergency Response Vehicles will increasingly be sent into environments that are dangerous for human beings, for instance, disaster or contaminated scenarios such a nuclear accident or unknown and unpredictable environments such as mining locations.

5.5.1.3 Infotainment

High resolution video streaming is nowadays a must in high-end vehicles and will even be more relevant in the future as self-driving cars become more common.

5.5.1.4 Vehicle management and diagnostics

This segment covers both fleet management and logistics activities, as well as tracking stolen vehicles and vehicle diagnostics and maintenance. Increasingly the latest cars today go on-line for remote diagnostics and maintenance. In the future this will become more widespread and in addition autonomous vehicles are likely to be fitted with black boxes, in the same way that planes are today. These must be protected for forensic diagnostic activity and to avoid insurance fraud. They must be capable of storing information at extremely high speed.
NGMN also lists two further use cases – nomadic nodes, where vehicles are used to extend the capacity of the network and assisted driving, which is an extension of vehicle use today where the driver receives information about traffic, driving conditions etc aimed at making driving easier.

5.5.2 Security requirements

The nature of the segment means that security will be critical in V2X as once again human lives may be at stake. 5G-PPP lists the main requirements as user authentication, authenticity of data, integrity of data, confidentiality, and user privacy [59].

Network coverage will be critical in many use cases in this segment for the vehicle must be always aware of its environment and should be able to interact at all time with it. Specific requirements will therefore include the ability to operate without network coverage, through non-3GPP means such as proximity services for device to device communications. Without coverage, the vehicle must still be able to receive and send information to other vehicles and pedestrians.

As the vehicles in any one scenario are unlikely to all be using the same network, lack of coverage for one network may lead to one vehicle with coverage acting as a relay device for another with none.

In all cases ultra-low latency is critical as a few milliseconds' delay could lead to a collision or a death.

Integrity, confidentiality, reliability and availability of the data sent is key even when there is no network coverage to be able to control that in the backend system. Mutual authentication is also critical to be certain that the data comes from the right/authorised entity and that it should be acted on. Low latency requirements extend to authentication to the network too.

Another big issue is user privacy, which is crucial to avoid tracking and the sharing of location and driving behaviours with unauthorised parties.

Overall, sector requirements closely echo many from enhanced mobile broadband and critical communications, for example the need to maintain communications while travelling at speed and consistent availability. The sectors even overlap, for example autonomous vehicles must be able to move out of the way of emergency vehicles. For some, V2X is a use case of critical communications rather than a segment in its own right.

[59] https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Automotive-Vertical-Sectors.pdf

5.5.3 Security requirements allocated by security layer

	Network	Service	Application	UE	Consumer
Protection of device identifiers	Integrity protection at rest and in transit; Identifiers' pseudonyms are used in transit		Privacy protection	Tamper-resistant storage of device ID	Privacy protection
Separation of device ID and subscriber ID			Application identities and credentials should be stored in tamper- resistant hardware	Separate tamper-resistant hardware storage means for device and subscriber IDs	Privacy protection
Secure storage of NAA				Requires the UE to provide a tamper-resistant entity	
Resist passive attacks	Forward secrecy				
No network coverage or nominal intra-vehicle communications			Mutual authentication between vehicles shall be possible with or without network coverage for either or both vehicles	Mutual authentication between vehicles shall be possible with or without network coverage for either or both vehicles	Privacy protection
Pre-provisioning during manufacture	No network access needed		Local provisioning during production could be seen as an application that is end-to-end protected	Vehicle connectivity credentials needs to be capable of being personalised during production - off-line personalisation Tamper-resistant hardware storage of credentials	
Secure access to remote provisioning system	Access to the provisioning system may be provided by either using the 3GPP or a non-3GPP network	Provisioning is a service provided on top of the network access. Access to the provisioning server needs to be secured			
Low latency	Optimised re-authentication; fast handling of security procedures	Additional service layer security may slow down processing	Additional application layer security may slow down processing	Dedicated crypto-HW to support fast processing of algorithms	
Geographical usage and location information	Localisation key to ensure correct data when roaming			Securely store the location information within the UE	Protect the location information for privacy reasons

Figure 5.9 V2X security requirements by security layer

This table shows security requirements allocated by security layer and reflects the fact that requirements mirror those of critical communications and enhanced mobile broadband.

5.5.4 Complimentary recommendations

The best solution to meet this segment's requirement for a highly robust level of security is secure tamper-resistant hardware. Given the ultra-low latency requirements, high speed tamper-resistant hardware with advanced crypto processing capabilities are required.

A range of security grades may be required. For example, police, military and government vehicles may be seen to require higher levels of protection. This can be done at the application layer, beyond network connectivity layers. Additionally, higher priority vehicle communications may be managed on a separate slice.

Any black box device must be capable of storing information securely and maintaining its integrity to avoid insurance fraud. Once again secure tamper-resistant hardware will provide the best approach.

6 Comparison between hardware and software approaches

We have seen throughout this paper that there are a highly diverse range of use cases and security requirements across each segment under consideration. What type of technology will best meet these requirements?

This chapter will consider three future alternatives in 5G to the removable UICC – eUICC or embedded UICC, Trusted Execution Environment (TEE) and Soft SIM. Please note that this analysis does not consider the Trusted Platform Module (TPM).

6.1 Definitions of each solution

The **eUICC** is defined by GSMA Intelligence[60] as "a UICC capable of supporting remote provisioning such as the GSMA Embedded SIM Specification."

The **TEE** is defined by GlobalPlatform[61] as "a secure area of the main processor in a smart phone (or any connected device). It ensures that sensitive data is stored, processed and protected in an isolated, trusted environment."

The **SoftSIM** is defined by GSMA Intelligence[62] as "a collection of software applications and data that perform all the functionality of a SIM card but does not reside in any kind of secure data storage. Instead it would be stored in the memory and processor of the communications device itself." It can also be viewed as an application executed by the application processor containing OS code, algorithms, keys (with / without white box crypto protection techniques).

Within the 5G environment, the SIM is likely to be embedded within the device in the form of an eUICC rather than continuing to be a removable SIM card as in today's mobile handset world. This has the effect of changing the business model of mobile service provision in some segments, but this document only considers those changes where they affect security.

Both SoftSIM and TEE as approaches see the UICC emulated outside of a secure tamper resistant hardware module. The difference between the two lies in where the emulated function is stored and handled.

In the latter it is stored within the TEE which is isolated from the Rich OS. In the former, it is stored in the Rich OS. As a result the TEE could be viewed as a SoftSIM operating within a TEE or a TEE based SIM.

Indeed, 5G technology is expected to be built around a "network of networks" concept and real-time handover between the network technologies involved will be key to success. NGMN identifies 50 Mbps throughput as an absolute minimum requirement (many envisioned applications require 1Gbps+) and the network will need to be optimised for data rate, latency, power efficiency and connection numbers.

Network operations also covers access to licensed or unlicensed, public or private networks. All these scenarios as well as the new features require security mechanisms that ensure authenticated access to the network both for regulatory reasons and for liability. They are applicable for all services utilising the 5G network.

60 https://www.gsmaintelligence.com/research/?file=81d866ecda8b80aa4642e06b877ec265&download

61 http://www.globalplatform.org/mediaguidetee.asp

62 https://www.gsmaintelligence.com/research/?file=81d866ecda8b80aa4642e06b877ec265&download

6.2 Pros and cons of each solution

6.2.1 eUICC

eUICC	
Pros	**Cons**
Security wise, it can be used for network layer protection across all use cases	For cheap devices, eUICC cost may be seen as a barrier
It is based on a mature solution (the UICC)	Soldered eUICC reduces flexibility
Lower cost for OEMs (connector space > free space + eUICC)	Testing / repair can be complicated (currently there is no test SIM support for testing)
It removes logistics costs of physical SIM distribution	Takes space on the PCB (vs TEE / SoftSIM although less than iUICC)
It allows industry preferred interface support (SPI / I2C) vs UICC (ISO)	If the eUICC fails, the device replacement is necessary
Isolated certifiable piece of hardware (via a protection profile) [this isn't possible with iUICC as it uses shared flash]	
VM interoperability at byte-code level allows operators to download the same app to any eUICC	

Figure 6.1 eUICC pros and cons

6.2.2 TEE

TEE	
Pros	**Cons**
Cost efficient if available on the chipset selected for the use case	Higher latency vs eUICC, which will be problematic for critical communications
Ease of integration	Compilation is needed for each type of TEE (native code)
TEE compliance testing (for the TEE, not the SIM) available at GlobalPlatform	SIM in form of a trusted application running on a TEE is not specified
Integration is extremely deep in the chip so there's a high speed interface (running at mb/sec)	Not yet proven for low-level authentication as it's not integrated into the baseband and hence means additional latency to gain access to resources
Certain application layer use cases are supported which are not possible on the eUICC side due to more processing power in a secure environment, with tighter integration (trusted UI)	Less secure as authentication mechanisms require dedicated non-Volatile Memory which the TEE currently lacks
TEE Firmware and application upgrade is easier because TEE is separate from Rich OS	Replay attacks leveraging design weaknesses above can lead to data integrity attacks which can lead to access to operator credentials which further can lead to massive denial of service attacks given the increased attack surface of the 5G mIOT segment
	Poor protection against physical attacks, as software is running in the external RAM of the solution, while secure element apps are running in embedded secure Flash
	If TEE fails, the device replacement is necessary

Figure 6.2 TEE pros and cons

6.2.3 SoftSIM

SoftSIM	
Pros	**Cons**
Saves space on the PCB inside devices, given that the chipset selected for the UE is powerful enough to process the SoftSIM	*Slow response time vs eUICC*
No cost to OEM	*Not yet proven for low-level authentication as it's not integrated into the baseband and hence means additional latency to gain access to resources*
No physical integration needed	*One version of each soft SIM per OS (and potentially per device variant)*
Simplified deployment model is possible via an app store for the eMBB segment for example	*Needs to be (re-)validated with each Rich OS or baseband firmware upgrade*
Power efficient (only running when needed)	*No clear security certification / audit scope possible due to software only architecture*
	Nothing standardised. May be device pre-requisites. A certain level of intra-vendor interoperability would be required
	Security based only on software. Any software based countermeasures consume memory and processing power which not all 5G segments can afford (e.g. IoT sensors)
	No means by which to ensure that sensitive assets are secured by a minimum level of software-based protection
	Poor protection against physical attacks, as software is running in the external RAM of the solution, while secure element apps are running in embedded secure Flash
	If software SIM fails, a fallback bearer would be required and not always present in every type of 5G Device (leading to device replacement)

Figure 6.3 SoftSIM pros and cons

6.2.4 Discussion of pros and cons

Clearly each potential solution for secure storage has pros and cons. The standout argument for the eUICC is that, based on mature technology, it provides a proven higher level of security than either the TEE or SoftSIM across all use cases. The major perceived negative is that it costs more, although that has proven not to always be the case for all ecosystem participants. In addition, throughout this paper, SIMalliance has urged that actors consider not only upfront costs compared to device costs when it comes to choosing security. Overall potential losses must be considered, including hidden costs that will only become apparent later.

Initially both the TEE and SoftSIM appear attractive for cost reasons because no hardware is involved. However, the SoftSIM will result in rising costs for certain actors, in particular MNOs. Set against this for the TEE is the significant disadvantage in certain segments that the emulation processing involved means that latency will rise, a disadvantage in critical communications, enhanced mobile broadband and V2X. Network authorisation time between different entities will become slower and may not be fit for purpose.

The TEE also lacks dedicated non-volatile memory, required by some authentication mechanisms for storing credentials. This greatly reduces security. While the TEE is a good solution for user authentication approaches such as FIDO [63], it cannot be recommended for low level authentication. One further significant disadvantage of the TEE is that the lack of non-volatile memory means that counters cannot be stored, opening up the way for replay attacks. As a result, despite its cost advantages and its ability to support Trusted User Interfaces, where a high level of security is required, it cannot be recommended as a replacement for a UICC or as an alternative to secure, tamper-resistant hardware storage.

[63] http://www.armtechforum.com.cn/2014/sz/A-8_FIDOandTEE-SimplerStrongerAuthentication.pdf

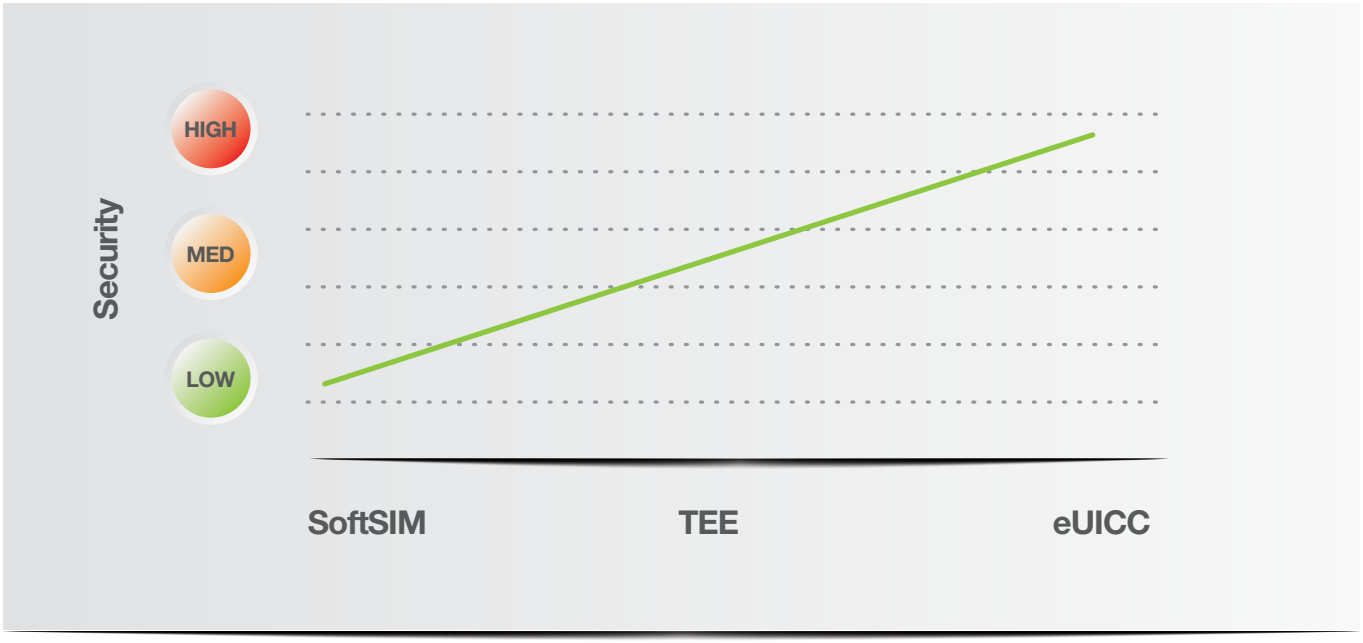

Figure 6.4 Security of SoftSIM, TEE and eUICC compared

The SoftSIM certainly appears attractive for cost (from the OEM perspective), power efficiency and ostensible simplicity reasons but in fact it too poses major disadvantages. Lack of standardisation at SoftSIM and device level results in a requirement for versions to be re-developed for each OS and device variant. Software only architecture makes security certification and audit problematic. At present the SoftSIM is not well defined enough to be taken seriously as a security option. There are no means by which to ensure that sensitive assets are secured by a minimum level of software-based protection.

It would also need to be (re-)validated with each Rich OS or baseband firmware upgrade. And once again, it poses latency issues due to lack of integration into the baseband, which make it not fit for purpose for critical communications, enhanced mobile broadband and V2X.

In addition, remote provisioning and management of the NAA need to be considered. While globally standardised interfaces and mechanisms exist for the eUICC, such mechanisms are not defined for SoftSIM and a TEE based SIM. This potentially leads to a variety of remote management or attestation mechanisms that are either chipset or OEM specific for both SoftSIMs and TEE based SIM. This increases complexity and consequently cost on the provisioning infrastructure of the MNO.

6.2.5 Solutions by use case

Based on these discussions, we can create the following suitability matrix, based on the current technical status of each solution. Data is classified by sensitivity. Ratings rank from '*' = low suitability to '***' = highly suitable. Scenarios left blank can be seen as 'not suitable', according to SIMalliance, but may be evaluated by the service provider on a case by case basis depending on data value.

- **Basic:** Machine information, unintelligible to an interceptor without location information.
- **Normal:** Machine or user data containing location information.
- **Critical:** Highly confidential / government / industrial secret user or machine information containing location.

Use case	Data sensitivity (impacts level of security required)	Solution Suitability		
		eUICC	TEE	SoftSIM
Massive IoT				
Home automation	Normal	***	*	-
Industrial critical	Critical	***	-	-
Industrial non-critical	Normal	***	*	-
Medical wearable	Critical	***	-	-
Consumer wearable	Normal	***	*	-
Metering	Critical	***	-	-
Retail stores / POS / Banking	Critical	***	-	-
Critical sensors	Critical	***	-	-
Non-critical sensors	Basic	** (as may not be favoured for cost reasons)	** (but TEE may not be present on such a low cost device)	** (as security protection may require more processing power and memory than SoftSIM can support)
Enhanced mobile broadband				
Laptop	Normal	***	- (for foreseeable future)	-
Broadband modem	Normal	***	*	-
Set top box	Normal	***	**	-
Smartphone	Normal	***	-	-
Tablet	Normal	***	*	-
Critical communications				
Critical communications including: drones, factory automation, e-health and public safety	Critical	***	-	-
VX2				
Automotive telematics	Critical	***	-	-
Automotive telematics	infotainment	***	*	-

Figure 6.5 Solution suitability matrix

This matrix suggests that for many use cases, especially where data is critical in nature, security requirements are such that the highest security tamper-resistant storage is clearly beneficial. As we have previously indicated, the level of security should match the value of the data being transmitted and decisions should be made on a case by case basis.

Bearing this point in mind, it seems from the comparison in this chapter that in the foreseeable future and certainly within the timespan envisioned for the launch of 5G, the eUICC can clearly meet those requirements while also taking into account functional requirements for speed and low latency too.

7 Conclusion

As this paper has shown, security requirements and challenges will be wider in 5G than in previous generations, reflecting the far broader range of potential use cases and potential threats.

Further contributing factors will come from the way 5G meets the need for higher speeds/lower latency combined with power efficiency needs, a wider variety of actors and device types and more use of the cloud and virtualisation.

The paper has surveyed use cases across 5 major segments – network operation, massive IoT, critical communications, enhanced mobile broadband and V2X. During that survey, security requirements have been identified and protective measures and mitigations recommended. These approaches will be summarised in this conclusion.

7.1 Summary of threats

This paper has discussed a wide range of threat types, including:

- Data manipulation
- Unprotected endpoint entry
- Equipment cloning
- Rogue devices
- Denial of service
- Eavesdropping
- Man in the middle attacks
- Impersonation attacks/spoofing
- Vehicle jacking/device theft
- Vehicle/device tracking
- Piracy of premium content
- Terrorist attacks
- Bidding down attacks.

Most of these attacks are applicable to all segments discussed to varying degrees. In order to counter these threats, SIMalliance believes that the following countermeasures are vital.

7.2 Key recommendations for securing 5G

Because network operations underlie all the other market segments, and in some cases accounts for most of their requirements, any 5G security recommendations must start with this segment.

Operators wish to reduce costs while expanding capacity, so it is highly likely that 5G will be built upon network slicing and the "network of networks" concept. Any security recommendations must take both this and mobile edge computing requirements into account. They must be able to work with 5G functional requirements for high speed, high reliability and low latency too.

Because each slice is a logical mapping of a set of functions, what is really important for security at this level is access control, authorisation and authentication (AAA) between individual virtualised functions. As a result, each virtualised function requires its own authentication mechanism and no compromised slice or function should be able to impact others. Mission critical functions, for example concerning subscription management or network authentication, must not be shared across slices.

Each subscription must be protected by a NAA within the UE that takes care of network identification, authentication and encryption. Each of these NAAs has its own identities and keys. Multiple NAAs may be active concurrently. Authentication may take place both at a network and service or application level.

For privacy, temporary subscriber identifiers are needed as are separate, independent key(s) for encryption and integrity protection. Keys for the four different security layers outlined in chapter 4 should be independent of each other. It must also be possible to securely store both the device identifier and the NAA separately from each other.

All aspects of device identification, NAA / subscription identification, authentication, integrity protection and encryption protocols should be optimised as much as possible to minimise the network attach and communication overheads without compromising the security.

At the same time, devices also need secure storage for application level credentials. In some sectors this storage may also need to store not just AKA authentication algorithms and keys but also algorithms and keys for non-3GPP RATs. In massive IoT it must also be able to cope with low power consumption requirements.

To take account of these requirements, any solution must have the following characteristics:

- It must meet functional requirements for high speed, high reliability and low latency and low power consumption.
- At the same time, it must offer highly secure tamper-resistant storage, with the ability to enforce rigid separation, of multiple keys, credentials and identities. This storage and its corresponding system should be certified by a third party.
- It must offer integrity and confidentiality (and in some areas privacy) protection.
- This solution must be capable of firmware upgrades and have a potential lifespan of as long as 15 years from deployment in some segments.
- It must offer a high level of trust in segments where errors may cause human lives to be lost.
- It must also offer the flexibility to allow for exceptions, in particular in critical communications where occasionally ultra-low latency requirements may over-ride other considerations.

SIMalliance considered three possible technical solutions to this list of requirements in chapter six. It concluded that, in the foreseeable future and certainly within the timespan envisioned for the launch of 5G, in uses cases involving high value data or human lives only the eUICC can meet every requirement while also taking into account functional requirements for speed and low latency too. In less high value use cases, other solutions may however also be suitable.

SIMalliance strongly urges that when it comes to making decisions about securing the future of 5G, security and risk requirements should take precedence over short term considerations about the bill of materials of an individual device.

Furthermore, when it comes to considering cost, the value of the data and the level of risk involved must be added to the question of device cost to obtain a full picture of the security needed.

SIMalliance believes that hidden costs that will become apparent over the next five years or more are almost a certainty.

That means that if security is not selected carefully today, based on use-case and alignment with the value of the data in transit, then it could cause problems tomorrow.

8 Appendices

8.1 Acronyms

5G	3GPP's Next Generation System
AKA	Authentication and key agreement used in 3G and 4G, provides guarantee to subscriber that they are connected to an authorised network entity
C-RAN	Cloud or Centralised Radio Access Network
eUICC	Embedded UICC
GP SEAC	GlobalPlatform Secure Element Access Control
HSS	Home Subscriber Server
IMEI	International Mobile Equipment Identity – the device identity in 2G, 3G and 4G
IMSI	International Mobile Subscriber Identity – the user identity in 2G, 3G and 4G
iUICC	Integrated UICC (integrated at chipset level)
LPA	Local Profile Assistant
LPWA	Low Power Wide Area
MCPTT	Mission Critical Push to Talk
NAA	Network Authentication Application
NFV	Network Functions Virtualisation
PCB	Printed Circuit Board
RAN	Radio Access Network
SDN	Software Defined Networking
TMSI	Temporary Mobile Subscriber Identity
UE	User equipment
UICC	Universal Integrated Circuit Card
USAT	USIM Application Toolkit
V2X	Vehicle to X

8 Appendices

8.2 Definitions

Backhaul	**The intermediate links between the core network and the sub-networks at the edge of the core network**
Control plane	**The control plane of a network carries network signalling traffic**
eUICC	**a UICC capable of supporting remote provisioning such as the GSMA Embedded SIM Specification**
GlobalPlatform	**GlobalPlatform is the industry body that defines and develops specifications to facilitate the secure deployment and management of multiple embedded applications on secure chip technology. Its work has enabled the logical partition of the SE into security domains, allowing the hosting of multiple credentials**
Network Authentication Application	**Application resident in UE takes care of network identification, authentication and encryption**
Soft SIM	**A collection of software applications and data that perform all the functionality of a SIM card but does not reside in any kind of secure data storage. Instead it would be stored in the memory and processor of the communications device itself**
TEE	**A secure area of the main processor in a smart phone (or any connected device). It ensures that sensitive data is stored, processed and protected in an isolated, trusted environment**
User plane	**The user plane of a network carries network user traffic**

5G Security – Making the Right Choice to Match your Needs

This paper was created by the SIMalliance 5G Working Group, which is made up of members from Comprion, Eastcompeace, Gemalto, Giesecke & Devrient, OASIS Smart Sim, Oberthur Technologies, Safran Identity and Security and VALID.

Special thanks to:

- *Paul Bradley (chair of the 5GWG)*
- *Patrice Beaudou*
- *Xavier Berard*
- *Elodie Clement*
- *Remy Cricco*
- *Claus Dietze*
- *Stephane Jacquelin*
- *Daniela Lopez*
- *Mireille Pauliac*
- *Stephan Spitz*
- *Eric Thérèné*
- *Jean-Christophe Vinatier*
- *Dragan Vujcic*
- *Jane Adams, iseepr (writer)*

About SIMalliance (Security, Identity, Mobility)

SIMalliance is the global, non-profit industry association which simplifies aspects of hardware-based device security to drive the creation, deployment and management of secure mobile services. The organisation promotes the essential role of a dedicated tamper-resistant hardware module in delivering secure mobile applications and services across all devices that can access wireless networks. By identifying and addressing related technical issues, and both clarifying and recommending existing technical standards relevant to the implementation of hardware security, the SIMalliance aims to facilitate and accelerate delivery of secure mobile applications globally.

SIMalliance members represent approximately 90% of the global SIM card market. As such, the SIMalliance's membership is responsible for delivering the most widely distributed secure application delivery platform in the world (UICC/SIM/USIM).

SIMalliance members are Card Centric Solutions, Eastcompeace, Gemalto, Giesecke & Devrient, Incard, Kona I, Oasis Smart SIM, Oberthur Technologies, Safran Identity and Security, VALID, Watchdata, Wuhan Tianyu and XH Smartcard (Zhuhai) Co. Ltd.

SIMalliance Strategic Partners are Comprion, Linxens and Movenda.