

5G: Towards secure ubiquitous connectivity beyond 2020

SICS technical report T2015:08

Martin Svensson

SICS Swedish ICT AB, Security Lab

Ideon Science Park
Building Beta 2 3v
Scheelevägen 17, Lund, Sweden

martin.svensson@sics.se
[@marsvesec](https://twitter.com/marsvesec)

Nicolae Paladi

SICS Swedish ICT AB, Security Lab

Isafjordsgatan 22, Kista, Sweden

nicolae@sics.se

Rosario Giustolisi

SICS Swedish ICT AB, Security Lab

Ideon Science Park
Building Beta 2 3v
Scheelevägen 17, Lund, Sweden

rosario.giustolisi@sics.se
[@saro_giu](https://twitter.com/saro_giu)



<https://www.sics.se/groups/security-lab-sec>

Dec 30, 2015

Abstract

The growing demand for mobile Internet, and the increasing number of connected devices, has required significant advancements in radio technology and networks compared to the previous generations of mobile telecommunication. Security however has only seen incremental changes to the previous mobile telecommunication generation, with enhancements that mitigate new threats and address revealed weaknesses. 5G is expected to change this, as novel use-cases will demand new trust models and require novel security solutions.

In this paper, we examine the state of 5G Security, and start by describing the new expectations, requirements and enablers in 5G and the design principles conferred by material presented in selected publications. Furthermore, we describe the historic development of the authentication and key agreement protocols, which were introduced with GSM (2G), as an example of the incremental improvements to security. Additionally, we present select published papers that suggest different types of attacks on the current generations of mobile networks, and solutions to the identified weaknesses, which must be taken into account in 5G security. Finally, we describe a proposed 5G Security architecture, which bring new models for authentication, authorization and accounting (AAA) to 5G.

The role of 5G security is clear, it must not only meet the basic security requirements in confidentiality, integrity and privacy, but also foster user confidence in mobile telecommunication.

Contents

Nomenclature	2
1 Introduction	7
2 Expectations, Use Cases, and Requirements for 5G	7
2.1 Expectations	8
2.2 Use Cases	9
2.2.1 Use case I: Internet of Things	9
2.2.2 Use case II: eHealth	10
2.2.3 Use case III: Safety-critical systems	10
2.3 Requirements	11
3 5G Technology enhancements and technology enablers	12
3.1 Radio network	12
3.1.1 Radio technology	12
3.1.2 Small cells	14
3.1.3 Supporting technologies	14
3.2 Core network	14
3.3 Trends and new business values	15
3.3.1 Plane changes (C/U) and slices	16
3.4 Driven by software	16
3.5 Design principles for 5G	17
4 Software Defined Networking Overview	19
4.1 Software-Defined Networking in Mobile Networks	21
5 Security	22
5.1 Background and evolution of GSM-LTE	23
5.1.1 GSM	23
5.1.2 UMTS (3G)	26
5.1.3 LTE (4G)	30
5.2 Weaknesses in UMTS and LTE	33
5.2.1 Authentication and Key Agreement protocol (AKA)	33
5.2.2 LTE Practical attacks	36
5.2.3 Transition to open protocols and hardware	38
5.3 Technology shift equals security shift	39
5.3.1 Security architecture	39
5.3.2 Software-defined networking	42
6 Conclusions	48

Nomenclature

1G	First Generation, page 21
2G	Second Generation, page 21
3G	Third Generation, page 7
3GPP	Third Generation Partnership Project, page 8
4G	Fourth Generation, page 7
5G	Fifth Generation, page 7
AAA	Authentication, Authorization, Accounting, page 37
AK	Anonymity Key, page 25
AKA	Authentication and Key Agreement, page 21
AMF	Authentication Management Field, page 25
API	Application Programming Interface, page 8
AS	Access Stratum, page 28
AuC	Authentication Centre, page 21
AUTN	Authentication Token, page 24
AV	Authentication Vector, page 28
BS	Base Station, page 31
BTS	Base Station Transceiver, page 21
BYOI	Bring-Your-Own-Identity, page 38
C-Plane	Control Plane, page 12
CK	Confidentiality Key, page 24
CN	Core Network, page 14
CS	Circuit Switched, page 14
D2D	Device-to-Device Communications, page 14
DNS	Domain Name System, page 36
DoS	Denial of Service, page 32
ECDH	Elliptic Curve Diffie Hellman, page 32

ECM	EPS Connection Management, page 32
EMM	EPS Mobility Management, page 35
eNB	Evolved NodeB, page 27
eNodeB	Evolved Node B, page 14
EPS	Evolved Packet System, page 27
ETSI	European Telecommunications Standard Institute, page 16
FDD	Frequency Divided Duplex, page 13
FIB	Forwarding Information Base, page 41
GN	Group Node, page 32
GPRS	General Packet Radio Service, page 22
GPS	Global Positioning System, page 15
GSM	Global System for Mobile Communications, page 7
GUTI	Globally Unique Temporary UE Identity, page 28
HE	Home Environment, page 25
HLR	Home Location Registry, page 28
HN	Home Network, page 28
HSS	Home Subscriber System, page 28
IK	Integrity Key, page 24
IMEI	International Mobile Equipment Identity, page 28
IMEISV	International Mobile Equipment Identity and Software Version, page 28
IMS	IP Multimedia CN Subsystem, page 36
IMSI	International Mobile Subscriber Identity, page 21
IoT	Internet of Things, page 9
IP	Internet Protocol, page 14
ISUP	ISDN User Part, page 36
ITU	International Telecommunications Union, page 8

KASME	Local master key in EPS, page 27
KDF	Key Derivation Function, page 29
KeNB	Intermediate key at eNB level, page 30
KNASENC	Key for NAS encryption, page 30
KNASINT	Key for NAS integrity, page 30
KPI	Key Performance Indicator, page 11
KRRCENC	Key for RRC encryption, page 30
KRRCINT	Key for RRC integrity, page 30
KSI	Key Set Identifier, page 24
KUPENC	Key for user-plane integrity, page 30
LTE	Long Term Evolution, page 7
M2M	Machine-to-Machine, page 37
MAC	Message Authentication Code, page 24
ME	Mobile Equipment, page 28
MEC	Mobile Edge Computing, page 16
METIS	Mobile and wireless communications Enablers for the Twenty-twenty Information Society, page 10
MIMO	Multiple Input Multiple Output, page 13
MITM	Man-In-The-Middle, page 21
MME	Mobility Management Entity, page 28
MS	Mobile Station, page 21
MVNO	Mobile Virtual Network Operator, page 37
NAS	Non-Access Stratum, page 28
NFV	Network Function Virtualization, page 7
NGMN	Next generation Mobile Network Alliance, page 8
OFDM	Orthogonal Frequency-Division Multiplexing, page 12
OS	Operating System, page 36

P-TMSI	Packet Temporary Mobile Subscriber Identity, page 32
PKI	Public Key Infrastructure, page 32
PLMN	Public Land Mobile Network, page 32
PPP	Public-Private Partnership, page 8
PTP	Precision Time Protocol, page 15
QoS	Quality of Service, page 15
RAN	Radio Access Network, page 12
RANAP	Radio Access Network Application Part, page 26
RAND	Random Number, page 23
RAT	Radio Access Technology, page 11
RES	Response, page 26
RNC	Radio Network Controller, page 24
RNTI	Radio Network Temporary Identities, page 32
RRC	Radio Resource Control, page 28
RSMA	Resource Spread Multiple Access, page 13
S-GW	Serving Gateway, page 14
SDN	Software Defined Networking, page 7
SE-AKA	A Secure and efficient group authentication and key agreement protocol for LTE networks, page 32
SGSN	Supporting GPRS Node, page 22
SIGTRAN	Signalling Transport, page 36
SIM	Subscriber Identity Module, page 10
SINR	Signal-to-Interference and Noise Ratio, page 13
SIP	Session Initialization Protocol, page 36
SLA	Service Level Agreement, page 19
SMComplete	Security Mode Complete, page 23
SN	Serving Network, page 24

SNid	Serving Network Identity, page 28
SQN	Sequence Number, page 23
SRES	Signed Response, page 23
SRNC	Serving Radio Network Controller, page 24
SS7	Signaling System No. 7, page 36
TA	Tracking Area, page 33
TAU	Tracking Area Update, page 34
TDD	Time Division Duplex, page 13
TMSI	Temporary Mobile Subscriber Identity, page 21
U-Plane	User Plane, page 12
UE	User Equipment, page 12
UMTS	Universal Mobile Telecommunication System, page 7
USIM	Universal Mobile Telecommunications System, page 32
VLR	Visitor Location Registry, page 22
VoLTE	Voice Over LTE, page 14
X2AP	Control plane protocol between eNodeBs and the X2 interface, page 36
XMAC	eXpected MAC, page 26
XOR	eXclusive OR, page 25
XRES	eXpected Response, page 26

1 Introduction

The growing demand for mobile Internet and increasing number of connected devices have introduced new capacity requirements for mobile telecommunications. Until today, the requirements have mainly been addressed with new physical radio transmission technologies that deliver higher bandwidth and lower latency. Despite the significant evolution of mobile Internet and major advancements in radio technology, the individual steps in security have often been based on incremental changes to the previous mobile telecommunication generation. New releases have brought security enhancements to mitigate new threats and to address revealed weaknesses. This is expected to change for 5G security. Novel use cases bring new types of requirements, in addition to bandwidth and latency improvements, hence 5G needs to be secured from its foundations.

We begin this report by summarizing published 5G visions, use cases and expectations in Section 2. These visions and use cases will in turn drive new requirements in the 5G architecture, requiring new enablers, which we describe in Section 3. In section 4 we focus on Software-defined Networking (SDN) and Network Functions Virtualization (NFV) and motivate their importance as enablers of 5G. We start the security aspects in Section 5 by describing the evolution of security considerations from GSM to the current version, Long Term Evolution (LTE, 4G). Considering that 5G security must take into account known weaknesses in universal mobile telecommunications system (UMTS, 3G) and LTE, we provide a description of select weaknesses and attacks in 5.2. We conclude the chapter with a proposed 5G security architecture from an active 5G research project. We conclude this report in Section 6 with a discussion and summary of some key points.

2 Expectations, Use Cases, and Requirements for 5G

The evolution of mobile telecommunication technology has been extraordinary in terms of available bandwidth and latency, which still remain important requirements for the development of new solutions. 3G brought integrated voice and mobile Internet, LTE drastically improved bandwidth and latency capabilities, and LTE Advanced have raised such capacity even further, producing the state-of-the-art of mobile telecommunication technology. While LTE is expected to support the needs of mobile telecommunications for many years to come, 5G will extend the support of devices over mobile telecommunications by building entirely new infrastructures consisting of heterogeneous technologies.

This section introduces the reader to expectations, use cases, and requirements for 5G, and is based on material presented in different publications

that address the topic.

2.1 Expectations

The expectation of 5G goes beyond the traditional definitions of consumer and operator as in today's networks. According to the white paper published by the Next Generation Mobile Networks Alliance (NGMN) 5G [1], 5G should support new value propositions and business models. For example, operator third-party partners should be able to access and control 5G services via application programming interfaces (API) that integrate well to the 5G system. Third-party partners and over-the-top players might address customers directly and offer services that are enriched by the operator network, connected *smart wearables* with remote monitoring being one example. It follows that security of telecommunications needs a careful reconsideration in terms of attacker and trust models due to the new involved parties. Moreover, the 5G network will have to be more flexible than today's networks. In fact, the 5G Infrastructure Public Private Partnerships (5G PPP) [2] envisions that 5G will be driven by software, in the context of software defined technologies such as NFV and SDN, to achieve the required design goals. Thus, the security of such software becomes of primary importance for the success of 5G.

Although 5G is still emerging as a technology, there are numerous organizations that have begun their research into 5G. The European Union has initiated 19 projects via the 5G Infrastructure Public Private Partnership (5G PPP), furthermore, worldwide 5G research is ongoing both in Asia and North America. The timeline for 5G – depicted in Figure 1 – suggests a first commercial deployment in 2020. We observe that each organization has specified different phases. The 3rd generation partnership project (3GPP) expresses the phases by release numbers. ITU names its phases with the expected contribution. 5G PPP foresees three phases:

- Phase 1 consists on specification of requirements;
- Phase 2 details research and optimization;
- Phase 3 is about experimentation and trials.

Although each organization assigns its own nomenclature for each phase, the different phases tend to synchronize among the organizations. In particular, the development of 4G and 5G in 3GPP is expected to synchronize with release 14 (R14) between 2016 and 2017 [2]. Still, some differences exist. However, to avoid that research in 5G takes very different directions among the organizations, a multilateral memorandum of understanding for “Global 5G Events” was recently signed between 5G organizations in Eu-

rope, USA, Japan, South Korea and China¹, which will hopefully ensure the continued synchronization of 5G research.

	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
5G 3GPP			R14	R15	R16					
4G 3GPP	R12	R13	R14	R15	R16					
ITU	Vision			Wkp	Proposals	Evaluation	IMT-2020 specifications			
5G PPP	5G PPP set-up	5G PPP Phase 1	5G PPP Phase 2	5G PPP Phase 3						
Summary Mobile Networks		Radio experiments			Trials		5G Deployment and commercialization			

Figure 1: 5G Timeline.

2.2 Use Cases

The new use cases for 5G introduce a set of novel requirements and cover a wide range of devices beyond smartphones. As discussed later, this also leads to the need of a new security architecture capable to support such requirements. Below, we introduce some novel use cases for mobile telecommunications. Our choice is corroborated by other research in the field [1,2,3].

2.2.1 Use case I: Internet of Things

The number of Internet-connected devices is expected to increase substantially, thus introducing a wide set of novel requirements and characteristics [1]. The collection of devices (or “things”) – embedded with electronics, software, sensors and network connectivity – is called Internet of Things(IoT) [4]. Smart wearables, sensor networks, and mobile video surveillance are examples of IoT devices.

5G is expected to fully support the connectivity of IoT. According to the NGMN 5G white paper [1], this use case is described as “support for massive IoT supportability”. In the context of 5G use cases, IoT devices are characterized by

1. low-energy;
2. low-cost;
3. massive deployments (that are to be supported by the 5G network), both as an overall aggregate and within the same cell.

¹Leading 5G Visionary Organizations in Europe, USA, Japan, South Korea and China Sign Multi-Lateral Memorandum of Understanding for “Global 5G Events”, <https://5g-ppp.eu/>

Due to its increased breadth and depth over existing network connectivity, IoT poses novel security challenges, such as authorization of SIMless as well as resource constraint devices. As 5G aims to be the network for excellence for IoT, it must provide an adequate level of security and reliability.

2.2.2 Use case II: eHealth

The term eHealth denotes the practice of supporting healthcare by electronic processes and communication. eHealth systems provide a win-win scenario for both patient and healthcare provider, because it allows the patients to remotely manage more of their own health care, and when necessary get remote assistance from healthcare professionals, something which may also reduce the costs for the provider.

To fulfill the high level of availability and guaranteed quality of service (QoS) required, as well as appropriate security levels to protect user privacy and confidentiality, eHealth demands ultra-reliable networks. Likewise, security requirements are amplified when sensitive personally identifiable information (PII) are exposed to public networks, such as the Internet. Meeting basic integrity, confidentiality, and privacy requirements is therefore necessary in order to ensure the trustworthiness of any eHealth service. According to a survey conducted by *The Economist* [5], 42% of the respondents in the public sector see the need to ensure patient privacy as the biggest challenge for letting the health industry adopt mobile health technologies.

The role of 5G security is clear. It has to contribute security to foster users' confidence on adoption of mobile health technologies.

2.2.3 Use case III: Safety-critical systems

Safety is an emergent property in computer systems, and connected devices are often placed in control situations within safety-critical systems. The area of safety-critical systems is an emerging market that demands reliable communications. The automotive sector is expected to be an important stakeholder for 5G communication, and the Mobile and wireless communications Enablers for the Twenty-twenty Information Society (METIS), an EU-funded project, have presented use cases for traffic safety, which include cars detecting safety critical situations — such as hazardous road conditions and accidents within reach of the car [6].

Safety-critical systems pose important challenges for 5G as their (security) failure could result in loss of life, significant property damage, or environmental damage.

2.3 Requirements

The use cases presented in [1, 2, 3, 7, 8] – in addition to the ones outlined above – will introduce a new set of requirements, beyond the traditional high bandwidth and low latency requirements. The METIS project [6] delivered scenarios, Key Performance Indicators (KPIs), and corresponding requirements for 5G mobile and wireless systems. It introduces five scenarios based on five challenges, presented in Table 1.

Table 1: Scenarios and challenges for 5G mobile and wireless systems.

<i>Scenarios</i>	<i>Challenges</i>
Amazingly fast	Very high data rate
Great service in a crowd	Very dense crowds of users
Ubiquitous devices communicating	Very low energy, cost, and a massive number of devices
Best experience follows you	Mobility
Real-time and reliable communications	Very low latency

Based on these scenarios they present 12 test cases with corresponding KPIs:

- Traffic volume density;
- Experienced user throughput;
- Latency;
- Reliability;
- Availability and retainability;
- Energy consumption;
- Cost.

The METIS project additionally specify several KPIs that 5G is expected to support. Among others, they present:

- 500Mbit/s average user data rate;
- Density of up to 900Gbps/km²;
- Mobility of 500km/h;
- Latency less than 5ms in 99.999% of the transmissions;

- $0.01\mu\text{J}/\text{bit}$ for a data rate in the order of 1kbps.

Similar requirements are presented by the NGNM Alliance [1]. Surprisingly, none of the KPIs directly target security, despite mentioning industries that are strongly associated with security guarantees, e.g. healthcare and eHealth.

3 5G Technology enhancements and technology enablers

To meet the use-cases and requirements mentioned above, 5G will have to evolve in several key technologies, in addition to the radio access technology (RAT). Below, we focus on selected technology advancements presented by telecommunication manufacturers to foster the evolution of 5G.

3.1 Radio network

5G vendors and operators suggest to support 5G with a new RAT that will evolve in parallel with current LTE technologies, including parallel work items in 3GPP radio access network (RAN) working groups [7,8]. Flexibility and the possibility to further evolve the RAT with later technology introductions are seen as a prerequisite in the development of the 5G RAT [7]. We present a summary of the proposed requirements of radio network capacity in Table 2.

Table 2: Proposed requirements for radio network capacity

Peak data rate 10 GB/s	Number of devices $1\text{M}/\text{km}^2$	Latency 5ms	Mobility ≥ 500 km/h
Mobile data volume 10 Tb/s/km ²	IoT terminals ≥ 1 trillion	Reliability 99.999%	Outdoor location accuracy $\leq 1\text{m}$

3.1.1 Radio technology

Companies and organizations within 3GPP have reached a consensus that 5G will need to utilize new frequencies, including a spectrum up to 100GHz, to support the high capacity and low latency use cases. On a high level, the spectrum is intended to be utilized as follows:

- In the frequencies below 6GHz, macro and small cells will provide “low” band 5G, coexisting with current technologies (2-3G, LTE (4G)).

- In the “high” band, i.e. frequencies above 6GHz, small cells will be used to support very high data rates and short-range connectivity to enable the ultra-dense network scenario.

The low spectrum is essential for economical delivery of mobile services, hence the availability of low spectrum bands is a priority, in addition to increasing the efficiency below 1GHz [1].

In mobile telecommunication architecture, the different types of traffic are normally grouped in planes. The management plane is used for managing the network itself, the control plane (C-Plane) carries signaling traffic, while the user plane (U-plane) carries the network user traffic. In this section we focus on the C-plane and U-plane. While the difference is conceptual, each plane is often implemented in overlay networks that are independent of each other. To optimize the use of the different frequencies -- due to their different properties -- a more convenient split of control plane and user plane should be evaluated, considering different upload and download paths. The split of the planes would imply multi-site connectivity from a single user equipment (UE), decoupling system information delivery and data functionality from different nodes [7].

Along with a new spectrum, a new modulation for 5G RAT should be considered [7,8]. Some vendors consider the Orthogonal frequency-division multiplexing (OFDM) as the best modulation technology for mobile broadband [8]. It is currently used in LTE networks and was chosen due to its robustness to multi-path fading, interference, and suitability to digital signal processing techniques. In addition to OFDM, Resource spread multiple access (RSMA) waveforms might be considered as an enabler for low-power IoT devices in 5G networks, considering its advantages for uplink short data bursts [8]. Moreover, 5G RAT is expected to support both dynamic time division duplex (TDD) and frequency divided duplex (FDD) to enable future unified spectrum utilization [7].

Another challenge for 5G is improving the signal-to-interference and noise ratio (SINR). LTE introduced multiple input multiple output (MIMO) antenna technology – albeit with a limited number of antennas. Further technology enhancements, i.e. massive MIMO, are seen as one of the possible solutions. MIMO is especially relevant for higher frequency bands, with properties that can increase the capacity and lower energy consumption and interference. These properties will also favor the operators network planning [9]. To fulfill both capacity and coverage needs in 5G, there are proposals to shift to a “Beam-centric NX design”, i.e. the UE will be mobile between beams rather than between nodes [7].

From a security point of view, there has been some criticism that LTE is vulnerable to simple jamming techniques. One of the weaknesses concern control instructions, which is only 1 percent of the total signal, but is vital for synchronization, needed to send or receive data [10].

Lastly, the 5G vision for network transmissions calls for a flexible and ultra-light design, exposed in [7]. In fact, several design principles concern the limitation of mandatory network transmissions, use of well-confined transmissions in time and frequency, and avoidance of strict timing relations. In a nutshell, data capacity must be able to scale independently of system overhead.

3.1.2 Small cells

Current macro cells are not considered sufficient to support the high bandwidth and massive number of devices in the ultra-dense deployments expected in 5G [2], therefore 5G will need a larger deployment of small cell technology. This results in additional requirements concerning automated network organization, e.g. self-configuration, automatic neighbor relation and self-healing mechanisms [3]. As uplink and downlink connectivity might be split in 5G, traffic asymmetry will increase, hence traffic management will need additional requirements as well as traffic assignment between RANs.

3.1.3 Supporting technologies

While there are several projects and publications regarding the radio technology for 5G, there is less research in the area of supporting technologies. Supporting technologies is a technology concept to further enhance the network capacity by the introduction of new technologies, such as caching or opportunistic communication. The literature only discusses supporting technologies in general, still concluding that enhanced RAT and small cell deployment will be insufficient to support the proposed use cases for 5G.

One of the possible techniques proposed to offload the network is the opportunistic device-to-device communications (D2D), [1,3], which is being researched by the MOTO project².

3.2 Core network

The heterogeneous use cases and bandwidth expectations of 5G must be supported also by the core network (CN) and backhaul, to not restrict the capacity offered by the RAN to users. The CN must also embrace a more open network architecture to support new business cases and values by third-party services, as mentioned above, in addition to enhanced bandwidth and latency. In the section below we present select technologies as enablers for these requirements.

²FP7 MOTO Project Website: <http://www.fp7-moto.eu>

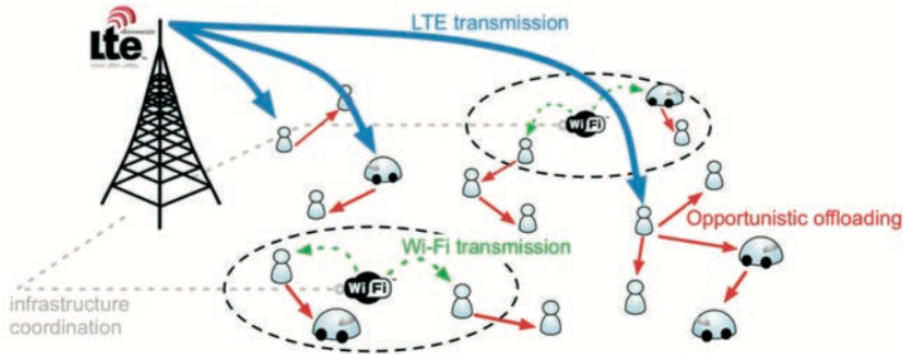


Figure 2: MOTO offloading techniques.

3.3 Trends and new business values

5G is expected to continue a trend started already with the introduction of LTE. LTE has converged to a less complex all IP-based network and a flatter architecture. In LTE, the user-plane consists of only two network elements: the Evolved Node B (eNodeB) and the serving gateway (S-GW), while the circuit-switched (CS) domain disappeared in favor of Voice over LTE (VoLTE). The eNodeB is a hardware connected to the mobile phone network, and communicates directly with the UEs. The S-GW transports the IP data traffic between the UE and the external networks, hence it deals with the user plane.

There are five major trends that the backhaul for 5G is expected to follow: open network architecture, end-to-end quality of service (QoS) and security, significantly higher data rates, reduced latency, and network-assisted synchronization [3]. An open network architecture is seen as a set of networks that are shared among operators. Virtualization will enable virtual sub-networks with network resources dynamically allocated among the operators, and neutral brokers that manage the distribution of resources that are priced according to offer and demand [3]. New use cases would make end-to-end QoS an essential enabler, suggesting that the RAN must actively verify the supported capacity in the backhaul via signaling and real-time QoS measurements to deliver guaranteed capacity [3].

Synchronization helps to mitigate inter-cell interference, thus increasing the spectral efficiency. To enable both indoor and outdoor installations, operators prefer network-assisted synchronization, a technology in which the backhaul assists with synchronization, over GPS synchronization. Network-assisted synchronization is currently based on two main approaches, IEEE 1588v2 precision time protocol (PTP) and Synchronous Ethernet, using the bit clock [11].

Related to capacity requirements, fog computing [12] is a technology

trend that is mentioned in the 5G PPP vision [2]. Fog computing is an architecture to move functionality, e.g. storage or communication configuration, closer to the edge of the network. It is seen as one of the key technological components to meet the performance targets.

The proposals mentioned aim to support new business cases and value creation for 5G. Thus, the 5G network will require flexibility and “as a service” approach in its design principles [2].

3.3.1 Plane changes (C/U) and slices

So far the CN has been seen as a monolithic design optimized for mobile broadband; instead, it needs to be rethought as a new infrastructure of heterogeneous technologies according to NGNM [1]. As anticipated in Section 3.1.1 – and suggested in the 3GPP RAN³ meeting – the control plane and user plane functions should be conveniently split to allow employing their functions on demand.

To support new models and flexible designs, the authors of [1] propose “5G Slice”, a network slicing technique to support the new use cases presented for 5G. Each slice consists of a number of network functions and RAT settings to support the specific use cases and business models of network service providers. A slice covers all parts of the network: software modules running on cloud nodes, configurations of the transport network, radio configuration, and the configuration of the 5G device itself. The goal of creating network slices is to provide exclusively the requested functionality. The request for a specific configuration is enabled by API calls to the 5G network.

3.4 Driven by software

The mobile telecommunication networks are already evolving towards an open architecture based on standard operating systems and hardware and is expected to continue towards a situation when “5G will be driven by software” [2]. Emerging technologies – such as SDN, NFV, mobile edge computing (MEC), and fog computing – are seen as enablers to achieve the performance and scalability goals. Work on NFV and SDN began in the European Telecommunications Standard Institute (ETSI) in 2012. At the SA#63 plenary 3GPP SA5 in March 2014, 3GPP began their work on defining NFV and SDN functionality in upcoming releases.

As mentioned above, the CN will be built as an open architecture with the ability for third-party organizations to request services directly from the network and to obtain QoS guarantees. SDN and NFV are expected to improve the flexibility of CN functions and of the allocation of resources.

³RAN 5G Workshop – “*The Start of Something*”: http://www.3gpp.org/news-events/3gpp-news/1734-ran_5g

Along with multiple virtualized components of the CN, many network parts will continue to run on dedicated, specialized hardware for performance reasons. Such hardware components should also be included in the SDN model, to allow the programmability of the C-plane [1].

Adoption of the SDN model for mobile networks leverages a series of advantages, namely: improved inter-cell interference management, improved mobility management, flexible support of virtual operators by partitioning flow space, distributed anchoring, and local break-out support and optimization of energy consumption.

NFV is complementary to SDN and has been enabled by advances in virtualization technology and hardware support for network processing, which allows to efficiently process network packages on commodity platforms. By implementing essential network functionalities – such as QoS monitoring, intrusion detection, firewalls, traffic shaping, etc. – in software applications on commodity platforms, NFV allows middleboxes to be removed from the network infrastructure.

Due to performance considerations, it would be naive to expect NFV to completely replace middleboxes from the network functionality. Rather, the two models coexist to better support the functionality of the network infrastructure. This is made possible, to a large extent, due to the abstractions introduced by network virtualization and the scalability of the management routines in the SDN model. Thus, identical configuration commands can be applied to all network management applications regardless of their deployment model – hardware, native, or virtualized.

NFV is currently in an earlier development stage than SDN, partly due to the lack of consensus over a so-called “Northern API”, i.e. the API between the network controller and network management functions. While there is a clean separation between control and data planes, the division between the functionality of the network controller and management applications is less clear. Thus, in many cases the applications are themselves a constituent part of the network controller.

We expect that the features of the SDN and NNF models will play an important role in the evolution of the next generation mobile telecommunication networks, by enabling new scenarios, such as support for the IoT devices, transient mobile network operators, and seamless integration with other enterprise networks. We describe the general architecture of SDN in more detail in Section 4.

3.5 Design principles for 5G

Table 3 summarizes the general design principles that we have presented in this section. The design principles are based upon the visions presented in [1, 2, 3, 7, 8].

Table 3: Summary of Design Goal Principles

<i>Radio Technology</i>	<i>Network</i>	<i>System Architecture</i>	<i>New Values</i>
Higher frequencies >6Ghz	Minimize number of entities and functionalities	Advanced Automation	Open interfaces – API
Unlicenced spectrum	C/U-function split	Built with modern OS architecture	Enable anything-as-a- service (XaaS)
Multiple connectivity	On-demand user-plane functions	Openness	Enhanced security
OFDM Modulation	RAT-agnostic core	NFV and SDN principles	
Massive MIMO / CoMP	Minimize Legacy	Network slicing	
Limitations of mangatory transmissions	Convergence between fixed and mobile services	Shared networks between operators	
Data capacity scaling independent of system overhead			
Small cell radio nodes (femto, micro, pico cells)			
Device to device communications			

4 Software Defined Networking Overview

The software-defined networking (SDN) model emerged and rapidly evolved in response to the increasing complexity of network deployments, allows facilitating operation and management of cloud-grade networks [13,14]. The operational advantages of the SDN model have led to its increasing adoption in enterprise-grade network deployments on a global scale [15].

A conceptual model of the SDN architecture is depicted in Figure 3 and described below based on the SDN architectural model presented in [16].

- The *data plane* contains both hardware and software routing equipment. This component implements the routing policies that fulfill the network administrator goals. It lacks decision logic and is optimized for forwarding speed. Packets that do not match any policy are either discarded or communicated to the control plane through the Southbound API.
- *Southbound API* is a vendor-agnostic set of instructions implemented by the routing equipment on the data plane. It allows bi-directional communication between the data and the control planes.
- *Control plane* is a logically distributed abstraction layer that transforms high-level network operator goals into discrete routing policies based on a global network view. It contains a distributed network operating system, which builds and maintains the global network view as well as communicates with the equipment on the data plane. The control plane also includes the network hypervisor, which multiplexes the available network resources among multiple users with distinct virtual network topologies.
- *Management applications* are used by network administrators to express their network configuration goals using a set of high-level comments. They could also include software-based network management components such as firewalls, intrusion detection systems, traffic shapers, etc.

In the process of operating the SDN deployment, the logically centralized control plane constructs a global view of the network components in its domain. This allows network management programs to rely on simpler graph processing algorithms to compute the shortest paths and to operate with higher-level abstractions, network operation is steered through network policies from three sources:

- High-level goals expressed by the network administrator and compiled into low-level configuration instructions for data-plane devices.

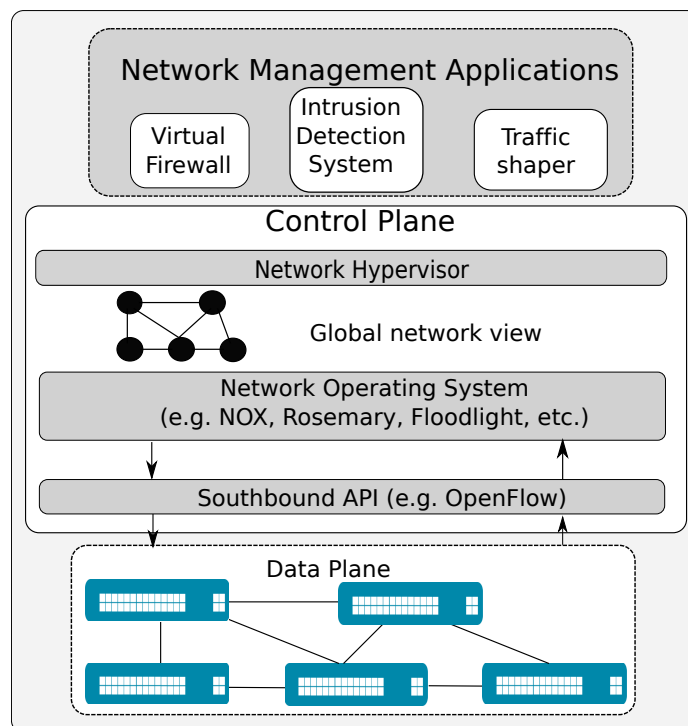


Figure 3: High Level Architecture of the SDN Model.

- Network management applications implemented as software components under the umbrella term network function virtualization (NFV) – which issue policies to implement their network functionality, e.g. as firewalls, traffic shapers, load balancers, intrusion detection devices and other functionality traditionally implemented in network middle-boxes.
- *Network operating systems* [17,18], which may independently generate network policies in order to ensure network liveness properties in the face of unexpected events (e.g. severe traffic anomalies or a DDoS attack on a subset of network components).

The continuous stream of policies from the sources described above – implemented by the network controller in a centralized manner throughout the deployment – leads to a continuous evolution of the network state. This introduces a new type of network configuration problems, since such network policies may have competing or conflicting effects on the data routing. In a security context, such network policy conflicts can lead to data leaks and isolation breaches in multi-tenant SDN environments. Thus, new algorithms are required for both static and run-time verification of network

configuration against policy invariants.

4.1 Software-Defined Networking in Mobile Networks

The Software-Defined Networking architectural model has so far received significant attention in the cloud computing context. This allowed to both evaluate its performance and evolve it according to the needs of both infrastructure and network service providers. Having been tested in large-scale enterprise deployments, the SDN model also applies to mobile telecommunication networks. Besides the generic advantages of SDN – such as improved ability to be managed, easier patching, flexible support of middleboxes – some additional aspects are specific for mobile networks [13], namely:

- *Better inter-cell interference management* – Centralized processing, inherent to the SDN model, allows implementing efficient radio resource management algorithms, in order to address the complex interference scenarios created by multi-cell interference. Furthermore, centralized processing allows – through inter-cell interference coordination – to improve performance by avoiding, canceling, or exploiting interference between adjacent cells. Finally, at the network level, centralized processing allows adding spectrum resources and configure the network to fine-tune user data traffic delivery, through orchestration and optimization of ultra-dense networks.
- *Improved mobility management*: An increase in the density of networks causes more frequent handovers due to the cell size. Thus, mobility management decisions become an important fact in mobility management, alongside with radio quality. Applying the SDN model can in this case shorten service disruption time and reduce switching costs, as well as enable effective load balancing.
- *Flexible support of virtual operators* by partitioning flow space – support for multi-tenant virtualized networks allows allocating, providing and enforcing network slice quotas according to the SLA agreed upon between the infrastructure providers and infrastructure tenants, i.e. network operators. Quota enforcement and efficient tenant isolation are key enablers of this use case and must be reliably implemented by a network hypervisor, or a similar component on the infrastructure control plane.
- *Distributed anchoring and local break-out support*: The centralized 3GPP network architectures create high traffic demands in the CN of network operators. Functionality supported by the SDN model can help mitigate this by distributing the user data plane, in order to allow local offloading of user data traffic. The decoupling between control

and data plane allows in this case to maintain a logically centralized control plane in order to enable globally optimized operation.

- *Energy optimization:* The global network view enabled by the SDN model allows the CN to optimize energy use by switching off parts of the RAN and backhaul – in order to reduce energy consumption – depending on user demand and network status.

Along with the rapid evolution of the state-of-the-art in network policy verification and enforcement for SDN deployments, as well as rapid progress towards mature and secure SDN controllers, a range of challenging problems and gaps continue to persist. Examples of such challenges are verifying liveness network properties (currently ignored in favor of safety properties), verifying policy composition for out-of-order rule installations, developing a model for non-interference among co-resident applications, as well as creating a sandboxing model for NFV applications interacting with the network operating system.

Similarly, a range of security risks – characteristic to SDN deployments – have been identified, such as vulnerabilities in the control plane, attacks in control plane communications, lack of a trust chain between the management applications and the data plane, attacks on policies and rules in programmable networks, resource limit violations, attacks on virtual switches and network gateways as well as weak bandwidth isolation as attack vehicle. We further expand on the above security risks in Section 5.3.

5 Security

Security has been an important part of the earlier success of mobile telecommunication – the public trust has been steadfast since the introduction of GSM to current LTE Advanced. One important aspect is that the security features have been transparent to the user and have been unobtrusive in its design, even between major versions, i.e. GSM to LTE.

As new versions have been released, new security functionality has been added to support both new business- and use-cases, and to mitigate identified weaknesses and attacks by enhancing the security protocols and security architecture. As this also applies to 5G, the security architecture and security protocols must be developed to support the novel use-cases and requirements expected in 5G.

In this section we begin with a historical review of the evolution of the authentication and key agreement protocol (AKA). This is one of the fundamental security protocols in mobile telecommunications and acts as a bootstrap protocol for communication. We describe the expanded use of confidentiality and integrity protection in each release to show how the AKA protocol has evolved to mitigate threats and known weaknesses. The section

continues with a presentation of select security weaknesses and attacks that have been exposed in UMTS and LTE. We conclude with a description of a proposed security architecture for 5G and the security aspects of SDN, expected to be prerequisite to enable the new business- and use-cases in 5G.

5.1 Background and evolution of GSM-LTE

Mobile telecommunication protocols have evolved with each new 3GPP release. The major releases, i.e. GSM to 3G to 4G/LTE, often receive the most attention with regard to higher bandwidth and lower latency, whereas the improved security mechanisms – though significant – are rarely mentioned. It is worth noting that the AKA security protocol was largely a success, and even though vulnerabilities have been identified, AKA is one of the most used security protocols in the world. In this section we review the historical development of the AKA protocol used in global system for mobile telecommunications (GSM), 3G and LTE, and its endpoint in the core network.

5.1.1 GSM

The requirement for GSM (2G) was to provide security on par with wired communications without loss of usability [19]. GSM moved to digital signaling from the analogue signals used in 1G, and brought new tools to increase security, such as cryptographic methods to protect the communication via authentication and confidentiality controls.

In a nutshell, the first version of the AKA protocol works as follows. The Mobile Station (MS) and the Authentication Centre (AuC) of the subscriber’s home network share a longterm secret key K_i ⁴ for each user i , stored in the AuC and the subscriber identity module (SIM) card. Authentication is performed by challenging the MS to perform a computation that is only possible with access to K_i ; authentication is successful provided the response is identical to the expected response, retrieved from the subscribers home network.

During the authentication process, a secret key K_C is established and used to confidentiality protect the communication between the MS and base station. Additionally, to improve user privacy, a temporary mobile subscriber identity (TMSI) is assigned to the MS as part of the initial signaling, to reduce the need to send the permanent international mobile subscriber identity (IMSI).

The design goal of the first version of the AKA protocol (GSM AKA) was to authenticate the mobile station and to provide session keys to confidentiality protect the wireless communications between the mobile station

⁴From UMTS and onward, the secret key was renamed to K , but to enhance readability we use K_i throughout the paper.

and base station. The GSM AKA procedure is described in Figure 4. Since the introduction of GSM, known weaknesses in the GSM AKA protocol have emerged into serious threats. One of the most discussed weaknesses is the fact that GSM AKA only provides a one-way authentication of the MS, and since there is no functionality for the MS to verify the base station transceiver (BTS), the protocol is vulnerable to false BTS attacks. A false BTS attack can enable an adversary to control all traffic passed via the air interface between the MS and BTS, i.e. Man-in-the-middle attack (MITM). The GSM AKA protocol also allowed an adversary to perform replay attacks by misusing previously exchanged messages. At the time of GSM AKA inception, it was considered too difficult for adversaries to build devices capable of transmitting GSM messages, hence no mitigations for active attacks - such as the attacks mentioned above - were included in the protocol. Additionally, as the security architecture was focused to make it on par with security on wired communications, which meant securing the air interface. That also meant that sensitive data, such as KC , was sent without protection within the networks – between the base stations, as well as between base station controllers and other nodes in the network. Another weakness is created by short ciphering keys, which made the protocol vulnerable for exhaustive search attacks with present computation capacity. The fact that the cryptographic algorithms used in the protocol were kept secret also started a public debate regarding its security.

GSM AKA detailed protocol description

Below is a detailed description of the GSM AKA protocol.

Authentication.

1. The AKA protocol is initialized by an “*Authentication data request*” from the MS, that includes the identity of the subscriber and the device capabilities, e.g. encryption algorithm support, that is sent to the Visitor Location Registry (VLR) (for the circuit-switched domain) or the serving GPRS support node (SGSN) (for the packet-switched domain). The VLR/SGSN hold a database of the subscribers that have roamed into its jurisdiction.
2. VLR or SGSN, depending on domain, prepares the challenge by acquiring authentication triplets from the AuC.
3. AuC, which holds the copy of the permanent secret for the specific IMSI, prepares the reply via the cryptographic functions A3 and A8
4. AuC responds with one or more authentication triplets consisting of a random number ($RAND$), a signed response ($SRES$), and the secret key KC to be used for confidentiality protection.

5. VLR/SGSN initiates the authentication of MS by sending the *RAND*, received from the AuC.
6. Upon receiving the challenge the MS prepares its *SRES** by using the same cryptographic functions as the AuC, namely A3, and also produce the secret key *KC* via A8.
7. MS sends the response *SRES** to the VLR/SGSN
8. VLR/SGSN compares *SRES** from the MS with *SRES* from the AuC; if *SRES** and *SRES* match, then MS is authenticated – concluding the authentication mechanism.

Encryption

9. VLR/SGSN sends *KC* to the base station transceiver (BTS), which is the network endpoint for encryption in GSM.
10. BTS prepares the encryption by selecting a cryptographic algorithm.
11. BTS informs MS of the chosen algorithm; MS is also assigned a temporary identity (TMSI), included in the response.
12. MS acknowledges the algorithm and TMSI with the *Security mode complete* (SMComplete) message, concluding the encryption establishment

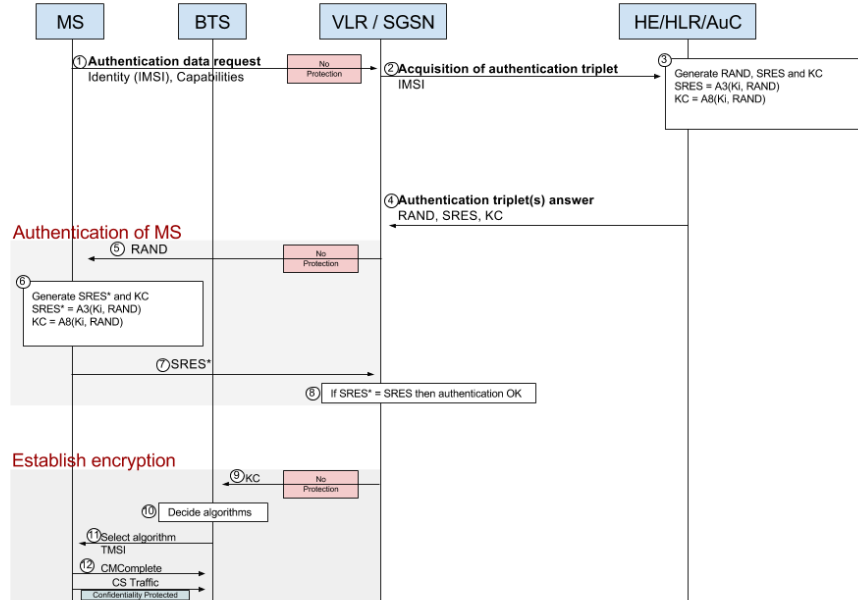


Figure 4: GSM AKA Procedure.

5.1.2 UMTS (3G)

UMTS maintained the principles introduced in the AKA protocol, however, the protocol was significantly extended with integrity protection and functions to prevent the possibility of replaying authentication messages. Replay attacks are effectively prevented by including a sequence number (SQN) in the challenge, and protecting the same challenge with a message authentication code (MAC). False base station attacks are also prevented with these changes to the AKA protocol, since UMTS AKA provides a mutual authentication which authenticates both the user and network. To maintain the public trust in UMTS, 3GPP decided to use publicly available cryptographic algorithms, as the secrecy of GSM cryptographic algorithms has earlier created controversial discussions. The cryptographic keys were extended to 128bits.

Similar to GSM AKA, the user equipment (UE) still authenticates with the VLR or SGSN via a challenge-response protocol. As mentioned, UMTS AKA includes enhancements compared to GSM, e.g. mutual authentication between the network and the UE and integrity protection of select protocols. The authentication vector is subsequently expanded with specific session keys for integrity (IK) and confidentiality (CK), an authentication token ($AUTN$) to enable the UE to verify the network and a MAC for integrity

protection. To mitigate replay attacks the *AUTN* contain a *SQN*. It is worth noting that since the *AUTN* is computed by the users home network, there is no possibility for the UE to authenticate the serving network (SN) in case of a roaming user. Instead, there is an implicit trust that the serving network is allowed by the UEs home network to provide mobile services, as the serving network is being able to retrieve authentication vectors from the home network. The UMTS AKA procedure is described in Figure 5.

In GSM, the circuit-switched confidentiality was terminated in the BTS, denoted NodeB in UMTS, which meant that all traffic from the BTS to the base station controller was sent unprotected, often via microwave link. To mitigate this security weakness, the cryptographic functions were extended to the serving radio network controller (SRNC) in UMTS, responsible for controlling the base stations, i.e. NodeBs, that are connected to it, which added integrity and confidentiality protection between the NodeB and core network, in addition to the higher level of physical security of the RNC compared to the base stations.

UMTS AKA detailed protocol description

Below is a detailed description of the UMTS AKA protocol.

Authentication.

1. The authentication and security mode setup is initialized by the connection establishment from the UE to the SRNC. Similar to GSM, the initial message includes the capabilities of the UE.
2. UE continues with the transmission of an “*Initial L3 Message*” to the VLR/SGSN, which includes the IMSI and a key set identifier (*KSI*). The *KSI* enables the re-use of the *CK* and *IK* during subsequent connections.
3. Similar to GSM AKA, the VLR/SGSN prepares the authentication challenge by requesting authentication vectors from the AuC in the home environment (HE).
4. AuC computes the authentication vector, which includes 128-bit integrity and confidentiality session keys, an *AUTN* to allow the UE verify the network, a *MAC* for integrity protection, the *RAND* and an expected response (*XRES*) that is similar to GSM. The *AUTN* consists of a *SQN*, which is optionally XORed with an anonymity key (*AK*) to conceal the *SQN*, an authentication management field (*AMF*) that can be used to control cryptographic functions and algorithms, and the *MAC*, i.e. $SQN \oplus AK \parallel AMF \parallel MAC$. The *MAC* is calculated with the cryptographic function *f1*, as $f1(Ki, AMF, SQN, RAND)$.
5. AuC sends the generated vectors to the VLR/SGSN

6. VLR/SGSN initiates the mutual authentication by sending the *RAND* and *AUTN* to the UE.
7. The same algorithms used by the AuC are applied in the UE to generate the necessary output, e.g. session keys, *SQN*, *RES*, expected MAC (*XMAC*) etc. With the successful execution of the cryptographic functions, the UE authenticates the network by validating that the *MAC* received in the *AUTN* is identical to the *XMAC*. The UE also verifies that the *SQN* is in the correct range, to prohibit replay attacks.
8. If the verifications succeed, the UE sends *RES* to the VLR/SGSN.
9. VLR/SGSN does a corresponding verification that the *RES* is identical to *XRES*; if true the mutual authentication is completed.

Encryption.

10. With identities of both the network and the UE verified, the VLR/SGSN initiates integrity and confidentiality protection by sending the RANAP message *Security mode command* to the SRNC, which is the termination point for the integrity and confidentiality protection in UMTS. The security mode command includes the allowed cryptographic algorithms for integrity and confidentiality protection and the associated session keys *CK* and *IK*.
11. SRNC decides the cryptographic algorithms based on a preference list from the VLR/SGSN and the capabilities sent by the UE in step 1. It generates the random number *FRESH*, and computes the integrity message *MAC-I* via a cryptographic function denoted *f₉*. The input for *MAC-I* is: the chosen algorithms; UE capabilities; the *FRESH*; a counter *COUNT-I*; a direction bit to indicate if the message is intended for uplink or downlink; the integrity key *IK*. By including the UE capabilities into *MAC-I*, a downgrade attack is effectively mitigated and the *COUNT-I* value protects against replay of earlier control messages.
12. SRNCs send the algorithms, UE capabilities, *MAC-I* and *FRESH* to the UE in a security mode command.
13. In a similar verification to the mutual authentication, the UE will compute its own *XMAC-I* and compare it with *MAC-I* to verify the integrity of the message. The UE will also verify that the received “*UE Security Capabilities*” are equal to the capabilities sent in step 1.
14. If verification is successful, the UE sends a “*Security mode complete*” message together with a *MAC-I* to the SRNC.
15. SRNC verifies the *MAC-I* with its own generated *XMAC-I*.

16. If the verification in step 15 succeed, the SRNC completes the security mode setup by sending Security mode complete – including the selected algorithms – to the VLR/SGSN.

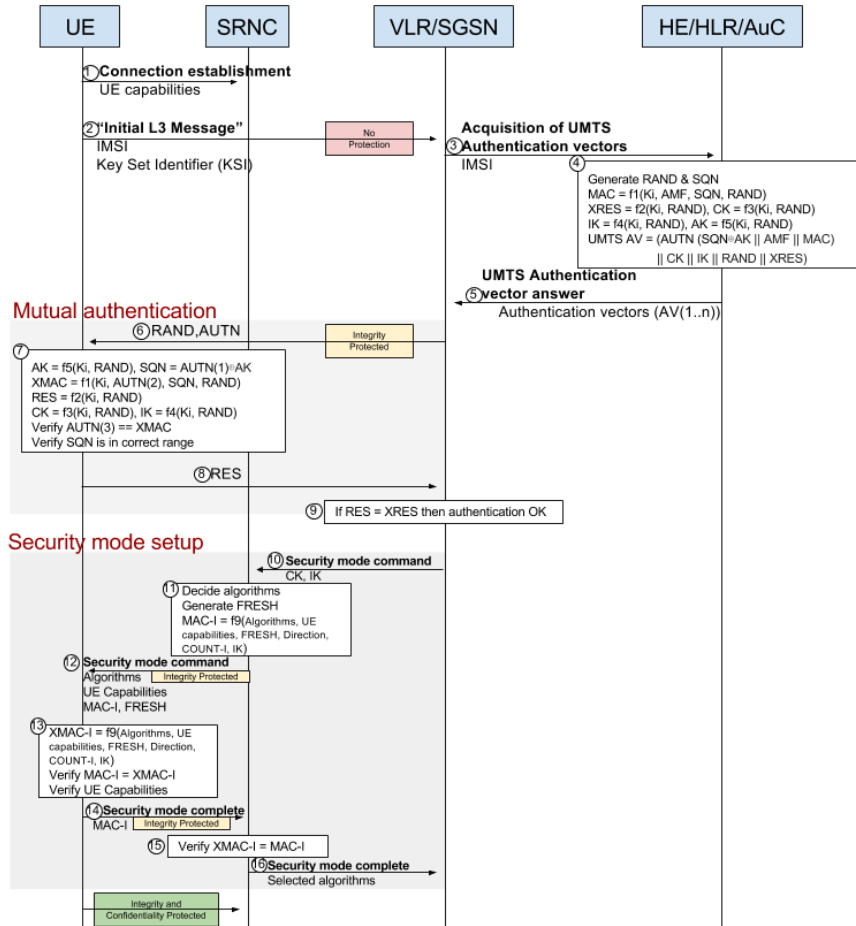


Figure 5: UMTS AKA Procedure.

5.1.3 LTE (4G)

Considering the success of UMTS security, 3GPP endeavored to alter it only where necessary to facilitate the new Evolved Packet System (EPS) architecture and the security requirements brought by changing business models or deployment requirements.

Thus, 3GPP continued the existing security association between the UE and AuC in LTE. With each of the sides storing and protecting the permanent key and thus continuing with the principles of AKA, the resulting protocols was denoted EPS-AKA. To prevent significant operator costs and to ease the consumer transition from UMTS to LTE (by avoiding the need to exchange their USIM), it was decided that LTE must support the UMTS USIMs. Due to major disadvantages of GSM AKA compared to EPS-AKA, the older GSM (2G) SIM cards were prohibited for LTE.

One of the design goals of LTE EPS was to flatten the architecture and discontinue the use of intermediate nodes, which made the base station – denoted as evolved Node B (abbreviated as eNodeB or eNB) in LTE networks – the termination point for many of the signaling protocols. This design restarted the discussion of the security termination point. Terminating signaling protocols in the eNodeB implied that the protection of those messages also terminated at the eNodeB. This is opposite to the decision of UMTS workgroup, which moved the termination point to the RNC – located deeper inside the UMTS network – to resolve the weakness of GSM. To mitigate the fact that the eNodeB was seen as unsecured (since it is placed in exposed locations), requirements were put in place to enhance the physical and system security of the eNodeB. This was the first time that 3GPP included specific platform security requirements for a network node. With these specifications in place, 3GPP accepted to terminate the security protocols in the eNodeB.

One of the high-level security requirements described in [TS22.278] is that a security lapse in one access technology must not compromise other accesses. Two significant changes compared to UMTS support this requirement. LTE introduced a distinction between the non-access stratum layer (NAS), handling traffic between the UE and core network, and the access stratum (AS) for signaling traffic between the UE and eNodeB. Additionally, LTE expanded the cryptographic key separation as a mechanism to limit the effect of a key leakage. In LTE a local master key, K_{ASME} , is derived from the UMTS integrity and confidentiality keys together with the identity of the serving network ($SNid$), which implies that the serving network is implicitly authorized as the HN has used the correct $SNid$ in its key calculation, which is an improvement from UMTS. From the local master key K_{ASME} , specific keys are derived to provide integrity and confidentiality protection for AS, NAS and RRC signaling traffic. The LTE AKA procedure is described in Figure 6.

Another feature of LTE is the improved privacy protection of the user, namely two specific changes in the handling of temporary identities as well as the permanent terminal identity of International Mobile Equipment Identity (IMEI) and International Mobile Equipment Identity and Software Version (IMEISV). To increase the privacy of the user, LTE has support for confidentiality protection of the signaling messages that transmit the Globally Unique Temporary UE Identity (GUTI) identity to the UE. If *used*, it prevents a passive adversary from correlating the GUTI identity with the permanent IMSI. However, active attacks to retrieve the IMSI from the UE are still possible. The second change is the required protection of the IMEI and IMEISV terminal identities, which are perhaps even more permanent than the IMSI since the user might switch operator more often than the mobile equipment (ME), by requiring NAS signaling protection before they are transmitted.

If we summarize the LTE advancements, the most significant changes were in the network part, making the design flatter and entirely removing the circuit switched domain. From a security perspective, the AKA protocol received improvements, enabling the UE to identify the serving network and the introduction of advanced key derivations.

LTE AKA detailed protocol description

Below is a detailed description of the LTE AKA protocol.

Authentication.

1. The authentication and security mode setup is initialized by the attach request from the UE to the MME, which is a central part responsible for paging, identity allocation, authentication among others in LTE.
2. The MME prepares the authentication challenge by requesting authentication vectors from the home subscriber system (HSS) in the home environment (HE) of the subscriber. The HSS contains user-related and subscription-related information and includes functionality such as user authentication and authorization, and is based on the home location registry (HLR) and AuC from earlier 3GPP releases.
3. Upon receiving the acquisition request from the MME, the AuC part of the HSS will compute an UMTS AV.
4. The HSS generates the extended authentication vectors, compared to UMTS, which include the local master key, K_{ASME} , calculated as $K_{ASME} = KDF(CK, IK, SNid, SQN \oplus AK)$.
5. The HSS sends the EPS authentication vector to the MME.
6. The MME initiates the mutual authentication by sending the $RAND$, $AUTN$ and a key set identifier (KSI_{ASME}) to the UE.

7. The same algorithms used by the HSS is applied in the UE to generate the necessary output, e.g. session keys, SQN , RES , $XMAC$, etc. The same verifications as in UMTS is performed by the UE to authenticate the network.
8. If the verifications are true, the UE send the RES to the MME.
9. The MME does a corresponding verification that the RES is identical to $XRES$, if true the mutual authentication is completed and the MME prepares the NAS security setup by deriving $K_{NAS_{ENC}}$ and $K_{NAS_{INT}}$ from K_{ASME} . The MME additionally produces non-access stratum MAC ($NAS-MAC$) used for integrity protection.

NAS Security setup.

10. The MME send the UE capabilities, NAS algorithms and the $NAS-MAC$ to the UE.
11. In a similar verification to the mutual authentication, the UE will derive $K_{NAS_{ENC}}$ and $K_{NAS_{INT}}$ from K_{ASME} and compute its own $XNAS-MAC$ and compare it with $NAS-MAC$ to verify the integrity of the message.
12. If verification is successful, the UE send an integrity and confidentiality protected NAS “*Security mode complete*” to the MME, concluding the NAS Security setup.

AS Security setup.

13. The MME derive a local eNodeB master key, K_{eNB} , from K_{ASME} .
14. The MME send K_{eNB} and the capabilities to the eNodeB.
15. The eNodeB derives cryptographic keys from the local eNodeB master key, K_{eNB} , to be used for encryption, $K_{RRC_{ENC}}$, and integrity protection, $K_{RRC_{INT}}$, of the Radio Resource Control (RRC) signaling protocol and user-plane encryption, $K_{UP_{ENC}}$.
16. The eNodeB initiates the AS Security setup, which includes the AS algorithms and $AS-MAC$.
17. The UE derives the AS session keys and the expected $AS-MAC$ and verify it with the $AS-MAC$.
18. If the verification is successful, the UE sends “*Security mode complete*” together with a MAC to the MME, concluding the AS Security setup.

authentication of the network, since only one key, KC , can be produced per authentication. This type of backward compatibility is exploited by several attacks.

Meyer and Wetzel [20] use this design decision to mount a MITM attack on the UMTS AKA protocol. The attack assumes that the adversary knows the victim's IMSI, which can easily be obtained by initiating an authentication procedure with the victim. With the IMSI known to the attacker, the attack consists of two phases. In phase 1 the attacker acts on behalf of the victim to retrieve a valid $AUTN$ from the real UMTS network. This is possible in UMTS since $AUTN$ and $RAND$ is sent without protection.

In phase 2, the attacker impersonates a valid GSM BS to the victim. Once the victim establishes a connection, it sends its security capabilities and TMSI, or IMSI, to the attacker. The attacker responds with the valid $AUTN$ and $RAND$ retrieved in phase 1, which the victim successfully verifies. In the subsequent security method setup the attacker decides to use "no encryption", or a broken version of the GSM algorithms. This attack succeeds if the time period between phase 1 and 2 is short so the SQN and $FRESH$ stays valid. The attack also requires that the victim's phone, denoted mobile station (MS) in GSM, allows roaming to GSM networks. With a successful attack the adversary is able to eavesdrop on all traffic between the victim and the mobile network.

Several contributions ([21, 22]) have focused on the fact that identities are sent without confidentiality protection during the initial authentication. Different proposals have been made to enhance the AKA protocol to provide confidentiality protection to IMSI and the temporary identities, i.e. TMSI, P-TMSI, GUTI and radio network temporary identities (RNTI).

The authors of [21] propose a new LTE protocol to protect the identities, such as IMSI and RNTI, from being transmitted without confidentiality protection over the air interfaces during the connection process. The threat model allows an attacker to track the victim using leaked identities, i.e. IMSI, and a number of rogue eNodeBs. This threat model also claims the possibility of LTE DoS attacks using the leaked identities. The authors propose a new protocol that includes a series of arithmetic operations together with exchanged random numbers and Public Land Mobile Network (PLMN) ID to generate keys, that are then used to safely transmit identities in the initial attach for UE in the ECM connection establishment.

Chengzhe et al. [22] propose a variant that addresses known weaknesses in EPS-AKA and also adds new functionalities. The proposed protocol is called SE-AKA: A secure and efficient group authentication and key agreement protocol for LTE networks. Due to the importance of backward compatibility, EPS-AKA inherits some weaknesses of UMTS-AKA, which the authors claim are addressed. SE-AKA introduce asymmetric key encryption and Elliptic Curve Diffie-Hellman (ECDH) which add properties of increased user privacy and perfect forward secrecy, a known deficiency in the current

EPS-AKA protocol. The protocol claims resistance to replay attacks, redirection attacks, MITM attacks and DoS attacks. The paper includes a formal verification of the security of the protocol by using ProVerif.

A lightweight public key infrastructure (PKI) is introduced in SE-AKA to provide each group node (GN) with a private/public key pair. The authors suggest to store the public key of HN in the trusted environment of the USIM. The public key enables the ME to encrypt the IMSI and therefore increase privacy properties of the protocol. Use of public key certificates to provide better protection for identities during the EPS-AKA design has been discussed within 3GPP; however, it was concluded that mandating a PKI infrastructure between all operators would be too costly [23].

SE-AKA also introduced support for group authentication. The emergence of machine-type communication has already begun with LTE and is expected to grow significantly, as mentioned in previous sections. To address the risk of high network access latency when numerous devices in a group need network access in a short timespan, SE-AKA introduces specific functionality for group authentication.

The weaknesses in EPS-AKA, identified in the presented papers, were often known already during the design of the protocols. In the published paper from JK. Tsay, and SF. Mjølsetnes [24] the authors present a vulnerability in both the UMTS-AKA and EPS-AKA protocols. The vulnerability exploits the fact that the SN has no means of associating an authentication data response from the HN to a specific UE, since its content is protected by the long term shared secret K that is not known to the SN. The authors present a scenario where the attacker can impersonate the victim to get a wireless service that will be billed to the victim by the HN. The attacker only needs the IMSI of the victim to initiate the attack, the victim does not need to be present on the network at the time. The attacker will initiate two concurrent AKA sessions to the SN, by sending both her own IMSI, and the IMSI' of the victim. As described in sections above, the SN will retrieve authentication vectors from the HN. During the subsequent execution of the AKA protocol, the attacker redirects the messages in such a way that they are interpreted by the SN to be intended for the victim's IMSI'. The real AKA session initiated for the victim's IMSI' is aborted, since the attacker does not have access to the secret key of the victim.

As mentioned in Section 5.1, the cryptographic protection in mobile telecommunication is based on a long-term shared secret K_i stored in the USIM of the user and the AuC of the network operator. If an adversary can get access to the shared secrets, she will be able to decrypt all traffic from the affected USIMs users. There are indications that this scenario has been exploited recently – a USIM vendor had a suspected breach of their security giving the attacker(s) access to the shared secrets of their produced

USIMs⁵. There are currently no mitigations available for this weakness – it is an intrinsic problem with all systems based on shared secrets.

5.2.2 LTE Practical attacks

Shaik et al. [25] describe the first practical attacks on LTE in their paper. The described persistent and silent attacks require user action – e.g. re-booting the device or have the USIM reinserted – in order for recovery. The authors describe 6 different attacks, 3 attacks which can lead to the exposure of the location (D1, D2, D3) of the target, and another 3 attacks that can cause persistent denial of service (L1, L2, L3). The protocols and signaling used in the location leak attacks are:

- Handling of identities and temporary identities on LTE networks;
- The paging and smart paging mechanism;
- The broadcasts and the information that is transmitted in these;
- The measurement reports that an UE can send to the network.

Paging is the method used to locate a specific UE in a particular tracking area (TA) to deliver a network service. A TA is a geographic area in a LTE network, which contains a group of cells. When a specific UE needs paging, the network broadcasts the page to the specific TA where the UE is registered. Additionally, a new functionality was added in LTE, called smart paging, that aims to reduce signaling overhead and to improve the time to locate an UE. In smart paging, only the specific eNodeB where the UE was last seen broadcasts the page. In the pages the UE is addressed by its temporary identity, if available, otherwise the permanent IMSI identity is used.

Shaik et al. have found a novel way to exploit these functions to locate the UE [25]. In the first exploit, L1, they utilize the fact that operators use the same GUTI identities for a significant time period, as long as the UE is powered on. Even if the UE is moved inside a city, the GUTI remains the same. Since the GUTI identity is persistent for several days it is possible to follow the subscriber’s movements. This attack is entirely passive, the attacker only needs to pick up the paging broadcasts.

The second attack, L2, is semi-passive. As mentioned in Section 5.1, the circuit-switched domain was removed in LTE and voice traffic transitioned to an IP-based solution, VoLTE. VoLTE has a high priority in LTE and therefore a TA wide broadcast will be used to quickly find the UE. By initiating a VoLTE call to the victim, long enough to cause a TA broadcast

⁵The Intercept: <https://theintercept.com/2015/02/19/great-sim-heist/>

but short enough to not trigger a notification on the victim's UE, the presence of a victim in a specific TA is found by observing the GUTI identities being broadcasted. By applying a method proposed by Kune et al. [26], the authors can identify the GUTI that is mapped to the phone number of the victim. When the TA is identified where the UE is located, the authors proceeded by using social networking applications to trigger a smart page. The authors have used specific *Facebook* and *WhatsApp* features to issue a smart page that do not lead to a notification on the UE. By using a smart page the authors could narrow down the location to a specific cell (roughly $2km_2$ in an urban setting).

The third location attack, L3, is an active attack using rogue eNodeBs. The attack is based on two different signaling methods, measurement report and RLF reports. These reports will contain the signal power from neighboring cells, and by using trilateration techniques the attacker can find the location of the UE. Although not supported by many UE vendors, LTE also has a function called 'locationInfo-r10' which will include GPS coordinates in the report, making the position even more exact. This measurement report attack vector is possible due to measurements reports are excepted from security requirements, i.e. sent in clear text without encryption and integrity verification. The protocols and signaling used in the denial of service attacks are:

- Tracking area update (TAU) procedure used to inform the network of the UE's present TA;
- LTE attach procedure that is sent unprotected and hence can be used for MITM attacks.

As with the vulnerability described in L3 above, the DoS attacks described below are possible due to the lack of integrity protection of RRC messages between the UE and eNodeB. The three vulnerabilities presented by Shaik et al. is based on two EMM protocol messages. The first is the TAU procedure used to update the MME of its current TA, and is used in the presented attack D1 and D2. The second procedure is in the "Attach" request which is used in the D3 attack. The attacks require the UE to connect to the attacker's rogue eNodeB.

In the first two attack scenarios presented by the authors, D1 and D2, the victim UE sends an integrity protected TAU Request to the rogue eNodeB, to which the attacker responds with a TAU Reject message. The LTE specification does not require encryption of these messages, hence it is possible by the attacker to send these to any of the UEs connected to the rogue eNodeB. The reject message can either degrade the UE service to UMTS or GSM, enabling further attacks, or deny all services to the UE. Regardless if it is a downgrade attack to UMTS/GSM or if all services are denied, the victims UE will stay in the new state until the UE is rebooted or if the

USIM is reinserted, thereby the denial of service attack can be considered persistent.

In the third denial of service attack, D3, the attackers perform a MITM attack during the LTE attach procedure. The attacker intercepts the “Attach request” sent by the UE to the eNodeB and modifies the message with the addition of “Additional update type – SMS only”, and forwards it to the eNodeB. The MME processes this message on behalf of the UE and executes the AKA procedure with the UE. The profile however only allows SMS and data services, as the attacker has denied voice services. Incoming or outgoing voice calls will be rejected by the MME. This attack is persistent until the UE is rebooted, the USIM is reinserted or if the UE is moved to another TA.

The authors in [25] present explanations and backgrounds as to why these vulnerabilities exist in their security analysis, they summarize them in three areas, security vs availability – security vs performance – security versus functionality. In all DoS cases (D1,D2,D3) and the third location leak, L3, 3GPP has documented the specification exception made from the security working group. If we look at the decision behind L3, as one example, the 3GPP security working group (SA3) suggested that all RRC protocol message should be encrypted, but in this case the availability aspect was considered more important than the privacy of the user. The reason behind the design decision was supposedly to enable measurement reports from all UEs, even if they are not able to establish a connection with the eNodeB and hence not being able to activate the security context. The paper summarizes it well in the sentence “*We show that the equilibrium points in the trade-offs have shifted today compared to where they were when the LTE security architecture was being designed*”.

5.2.3 Transition to open protocols and hardware

With the introduction of LTE a considerable architectural change was made as the circuit-switched domain was removed in favor of an entirely packet-switched IP-based architecture. Additionally, the telecom industry is moving from a monolithic design to off-the-shelf hardware and operating systems, such as Linux [21]. The protocols used in LTE has made a similar transition to utilize more open protocols, such as session initialization protocol (SIP) for VoLTE, Diameter for AAA, and domain name system (DNS) for the IP multimedia subsystem (IMS).

The presentation [27] at the HITB conference 2013 suggest that this transition will increase the attack vectors of LTE networks. VoLTE is based on SIP but with injection of signaling system #7 (SS7) via ISUP. DNS is used extensive in IMS, but without the required security it becomes an open directory for the entire network, including the location of equipment and identities of users, according to the author. Vulnerabilities and mis-

configurations that are common in traditional IT infrastructures, e.g. OS exploits, can now be used to attack LTE networks.

The presenter implies that instead of decreasing the complexity, by moving to a flatter design, IP-based infrastructure and the adoption of open protocols that are “wider”, the complexity will double as all the traditional protocols will still be required. SIGTRAN and SS7 will be tunneled via IMS and *Diameter*. An example of the increased network attack surface in LTE is the X2AP interface. In LTE, eNodeBs can address other eNodeBs directly, compared to 3G where the BS was only connected to the RNC. This results in a lower defense in depth security, if an attacker can gain access to one eNodeB the attacker will have layer 2 access to the other eNodeBs.

5.3 Technology shift equals security shift

In this section we describe a proposed AAA security architecture for 5G and how software defined networking and network virtualization functionality can enable the use cases.

5.3.1 Security architecture

In the evolution of mobile telecommunication networks, security has never been the main driver. That is not to say that the security has been at a stand still, quite the contrary, security protocols have been significantly enhanced from GSM 2G to LTE, as mentioned in previous sections. But security has previously been bolted to new technology enhancements and existing security mechanisms has been reapplied to new networks. For 5G this is no longer an option [28].

According to [2], there is a key difference with 5G in the targeted business and service delivery models. New actors bring new security requirements that needs to be handled as well as new types of interactions between actors. 5G PPP takes note that new services are expected to be deployed through virtualization to lower costs and to increase agility. A telecom server is no longer a special physical box with proprietary protocols that is located in a physically secured location, which introduce new attack vectors [27, 28].

To enable these new actors, services and business models 5G PPP call for “drastically new trust models and a security architecture built from the ground-up” [28].

The new trust model must take into account that, as has been known from the recent leaks, cyber-attacks are a real threat, both from states and individuals. Today’s trust model is based on a friendly environment between operators and connected devices. The new actors also alters the concept of “operators”: in 5G a car manufacturer might deploy 5G, hence the manufacturer becomes a mobile virtual network operator (MVNO). The aforementioned transition from dedicated telecommunication hardware to open

platforms, or even infrastructure-as-service or platform-as-a-service cloud service models, further reinforces the need for a new trust model. The existing trust models cover only parts of the 5G scenarios at best, additionally, as a consequence from the recent leaks, security is currently seen as a prerequisite when launching new services today, not only in 5G

In November 2015 a dedicated security project, 5G-ENSURE⁶, was started as part of 5G PPP. The project will include the development of a proposed security architecture and security enablers in the areas AAA, Privacy, Trust, Security Monitoring and Network Management & Virtualization isolation.

Existing AAA methods, e.g. EPS-AKA, must be adapted and be made more flexible to support new use cases, models that are bound to single user subscriptions must adapt to handle massive machine to machine (M2M) or IoT units. Identity bearers cannot be expected to be hardware security elements in all types of devices, car manufacturers present 'SIMless' communications as one of the use cases, trends and requirements in their presentation at the 3GPP RAN 5G workshop [29]. The challenge for 5G will be to support the extensive AAA legacy framework and still be able to add the required extensions by new business models and technology advancements.

In the 5G-ENSURE proposal [28], Figures 7 and 8 below are presented as an overview of the target security architecture and AAA scenario for 5G. The architecture in Figure 7 is based on ITU-T X.805. The foundation in AAA is the secure identification of the services, end-points and users. As mentioned above, 5G must support devices that lack a dedicated hardware module, i.e. USIM, that holds the identity of the user. As described in the 5G-ENSURE proposal [28] the existing binding between the USIM and HLR as the primary method to handle credentials will be far too expensive and inflexible when scaling up to massive number of devices expected in 5G. 5G-ENSURE will present new models, flexible and secure to add support for SIMless devices and with a distributed approach for authentication to support inter-operability. Additionally, the distributed approach might also involve the bring-your-own-identity (BYOI) concept. In many cases, enterprises already have their existing AAA infrastructure in place for employees and devices. In a BYOI configuration, these devices and users can re-use their pre-existing identities, e.g. certificates issued to devices, as a basis for 5G access.

The need for new methods also expand into authorization. In existing mobile telecommunication networks the authorization methods are not adapted to support resource constrained environments, both with regard to computing and network resources. In 5G new models are needed that provide granular access control decisions and are taken as close to the application as possible.

The future protocols can be based on existing protocols, e.g. OAuth

⁶5G-Ensure Project Website: <http://5gensure.eu>

and OpenID Connect, that are integrated with the 5G authorization infrastructure. These models, protocols and interfaces need to be standardized to prevent market fragmentation [28].

As stated in the proposal, accounting is mainly a business enabler, nevertheless 5G-ENSURE will work with accounting as an important enabler for 5G. In 5G, the user might take advantage of many kinds of services, which emphasize the need for assurance to the user regarding the cost of these services, and the resource usage. Conversely, the operators do not want the users to dispute the cost of the used services. Thus, new business models and new business actors require new secure and assured accounting principles, including non-repudiation properties.

In Figure 7, the model is showing both the logical and functional format. The functional dimension is represented by the yellow and blue boxes. The blue boxes illustrate the required security capabilities, while the yellow boxes represent *enablers* in each capability area. Together they comprise the functional security services for Data Security – Authentication – Authorization – Availability – Trust and Security monitoring and Privacy.

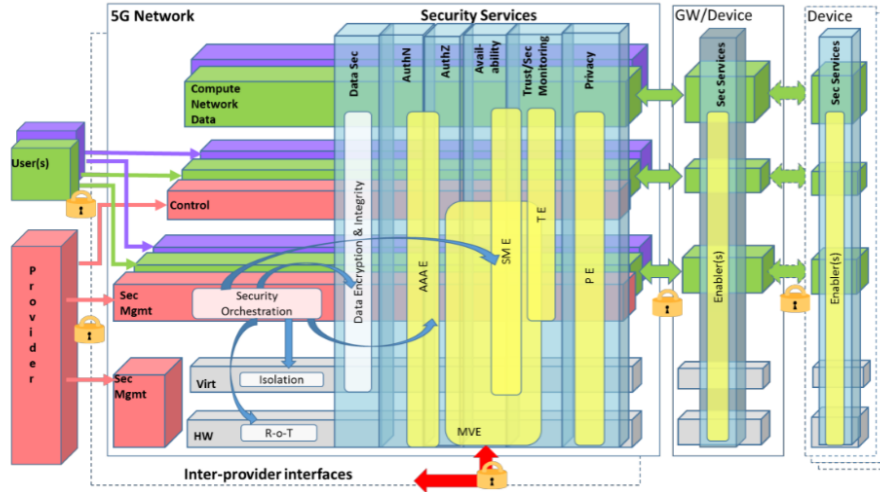


Figure 7: Proposed security architecture for 5G.

The logical dimension is shown by the horizontal boxes and illustrates the layering of the 5G network into planes, and the “slicing” of planes for tenants and providers.

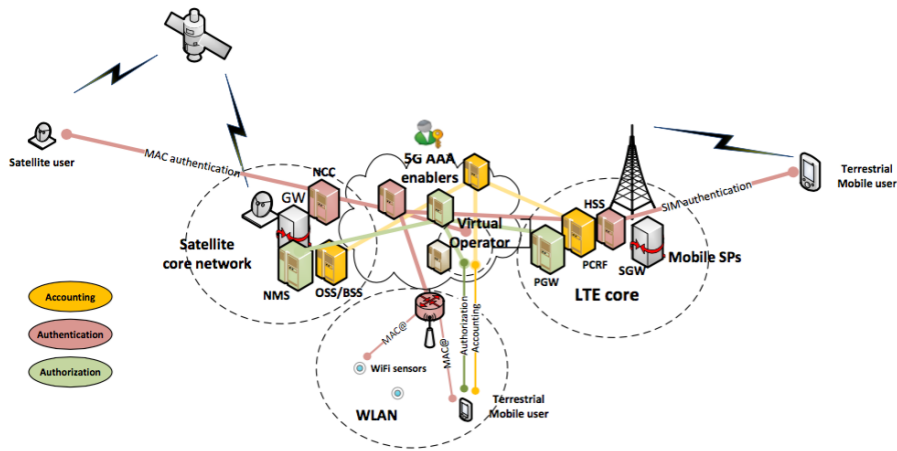


Figure 8: Proposed AAA architecture for 5G.

5.3.2 Software-defined networking

As mentioned earlier, SDN is expected to be a prerequisite for enabling the use cases of 5G. The SDN architectural approach challenges many of the network infrastructure rules and best practices that have evolved over the decades since packet-switched digital network communication gained popularity and to a large extent replaced circuit-switched networks. Likewise, many security best practices accumulated over the years are becoming increasingly obsolete and must be adapted to the new architectural model in order to adjust to the newly emerging risk factors and threat vectors.

One such risk factor is the centralized, global view of the network components (collectively called the network edge) and the links between them maintained by the network controller. Such functionality introduces multiple new capabilities for improved network management – anything from pre-calculating optimized traffic routing to software applications replacing hardware middleboxes. However, it also introduces a single point of failure – the network controller – which operates on a global network view built from on its recurrent communication with the SDN-enabled network components. Once compromised, the controller can provide the attacker with complete control over the entire network. An attacker capable of impersonating network components can thus distort the controller’s global network view and influence the network-wide routing policies.

Another potential risk factor is the proliferation of virtualized network components (such as routers and switches) running on full-fledged commodity OS, often assigned the same trust level and privileges as specialized, hardware network components with compact embedded software [30]. Considering that commodity OS with large code bases are likely to contain

multiple exploitable security flaws, such components can be attacked and modified to not follow the protocol, reroute traffic to a malicious destination or hijack other network edge components through lateral attacks.

Below we present an enumeration of the attack vectors applicable to the SDN model, based on [31]. Each of the presented attack vectors is accompanied by requirements that – if implemented – would help mitigate the risk introduced by the attack vectors. The full analysis can be found in [31].

Vulnerabilities in the control plane

Along with ease of network administration, a central control plane introduces a primary attack target for an adversary motivated to take control of the network. Taking over the control plane component in the SDN architecture allows the adversary to obtain full control of the network communication, different from traditional networks where communication control is distributed throughout various network components. Possible solutions include splitting the controller into several domains or distributing the control plane over several hosts, such that issued policies are verified on a different component before deployment.

- The SDN control plane must implement an access control model which limits the effects that vulnerabilities in controllers can have on tenant domains. This can prevent an adversary from simultaneously gaining control over the functionality of the SDN controller at all privilege levels and in all roles.
- A dedicated entity must verify the policies to be implemented by the SDN control plane before deployment.

Attacks on control plane communications

To manipulate network policies, the adversary may attempt to spoof the control plane communication (both among the components of a distributed controller and between management applications, controller and data plane). Similar attacks have been discussed in the context of mobile communication networks, where nodes had to be protected from spoofing attacks and redirect attacks [32], which allows researchers to build up on a rich body of knowledge in the field. However, several important differences must be noted: nodes in SDN deployments are static, which simplifies the task by ignoring any mobility or hand-over issues; nodes in SDN deployments are not constrained in terms of computational power or energy supply, which allows one to use the full range of cryptographic tools; finally, the network endpoints which are part of SDN deployments are fundamentally under the

control of the SDN network provider, which limits their ability to perform attacks on the network infrastructure.

When it comes to attacks on control plane communications in SDN deployments, possible solutions include enforcing authenticated and encrypted communication between all the control plane components, as well as secure enrollment mechanism for management applications and data plane devices.

- All communication between control plane components must be authenticated, and a secure enrollment mechanism for management applications and data plane devices must be in place.

Lack of a trust chain between the management applications and the data plane.

While the effort on defining the SDN architecture is still in progress, it is clear that management applications belong to a different security domain than the network operating system, and can be launched by malicious administrators or issue conflicting policies. Both detecting and preventing malicious policy deviations is challenging: a tenant can only observe the traffic after a change has been applied, but can not obtain and examine snapshots of the data plane forwarding information base (FIB); similarly, there is no mechanism to establish a trust chain between tenant commands and entries in the FIB. Possible solutions can be adapted from the ones employed – with varying success – on platform operating systems: verification of code origin and information flow control; however, such mechanisms do not satisfy malicious policy detection requirements.

- A mechanism must be in place to offer traceability and non-repudiation for all configuration commands and policies issued by network management applications.

Attacks on policies and rules in programmable networks.

Even if the integrity of policies remains intact, the adversary may issue malicious policies that modify or disable the effect of legitimate policies already in place (specifically in the scenario with such network management applications implement functionality of network middleboxes). This type of attack is difficult to detect and prevent, since the malicious policies might be indistinguishable from legitimate ones up to the point when the combined policy is deployed (furthermore, it requires a robust definition of a “malicious policy”). Possible solutions are to establish policy hierarchies and perform policy integration verification against some pre-determined invariants prior to deployment, to ensure that the resulting modifications remain within the basic policy framework. As policy updates may occur interactively in response to changing network patterns, both static analysis of policies and a pre-deployment simulation may be required.

Further, a set of existing challenges for policy verification and enforcement in SDN deployments are outlined in [33], such as verification of liveness properties, interleaved execution, determining verification time, enforcing non-interference, and finally sandboxing network applications.

We identify the following minimal principles to reduce the impact of attacks on policies and rules in programmable networks:

- A mechanism must be in place to enforce strong network policy isolation, such that the effects of policies in a certain tenant domain have no effect on other domains. Furthermore, the infrastructure provider must be able to enforce strict boundaries on the effects of policies within tenant domains.
- New network management policies must run through an integration verification engine prior to deployment, to minimize or exclude the effect of malicious policies on the network configuration.

Resource limit violation.

A malicious tenant may deploy network management applications that exploit vulnerabilities in network service isolation in order to gain network resources beyond the allocated quota defined in the QoS agreement. Possible solutions include adding network operating system capabilities for fine-grained monitoring of management applications to prevent resource over-allocation. This in turn requires a well-defined network resource model based on clear definitions of network resources and their respective capacities.

- A mechanism must be in place to ensure that network management applications do not allocate resources beyond the assigned quota. To do this, the NOS may apply advanced policing mechanisms – e.g. based on existing extensions, such as in [34] – that keep fine-grained tracking of management applications resource utilization and prevent them from making over-allocations.

Attacks on virtual switches and network gateways.

As pointed earlier, an adversary that controls a virtual network infrastructure component (such as a virtual switch) can attempt to impersonate other virtual network infrastructure components, spoof traffic and negatively affect tenant isolation. Possible solutions include integrity verification of virtual network infrastructure components and protecting the cryptographic secrets necessary for network access using a hardware root of trust.

- Integrity of virtual network components must be verified prior to deployment and the cryptographic material required for their network access must be protected with a hardware root of trust.

Weak bandwidth isolation as attack vehicle.

One of the consequences of NIC virtualization is a weakening of QoS guarantees, since most NIC virtualization implementations do not support guaranteed bandwidth [35]. While this does not directly affect data integrity and confidentiality, manipulating bandwidth allocation between tenants sharing a resource can be used in order to force a policy change (e.g. trigger a more permissive policy that is activated when the available bandwidth falls below a certain threshold). Possible solutions include widespread proliferation of bandwidth isolation techniques such as described in [36], as well as including the effects of bandwidth changes into network policy security testing.

- Policy-based routing decisions must not be affected by vulnerabilities in bandwidth isolation between tenants. To clarify, consider a network setup with two types of paths: low-bandwidth, low-cost, low-security permanent paths (type-A paths) and high-bandwidth, high-cost, high-security switched paths (type-B paths). Consider further that a legitimate tenant has configured a policy to distribute different types of traffic (low-value and high-value traffic) among the type-A and type-B paths respectively. An adversary capable of modifying the bandwidth allocated to the paths of the legitimate tenant should not succeed in redirecting high-value traffic through type-A paths.
- Software and hardware network components must offer equally strong bandwidth isolation properties. In the current networks, the data plane components include both software switches and routers deployed on commodity platforms and specialized hardware equipment implemented with application-specific integrated circuits. As pointed out in [23], software-based data plane components lack many of the features currently implemented in specialized hardware switches and routers. Strong bandwidth isolation is one of the features which must be improved in the software implementations.

Information leakage between network slices.

The risk of leaking information through side-channel attacks is another aspect that emphasizes the importance of isolation between network slices. Collocated virtual network tenants sharing common physical infrastructure may infer information by analyzing observed traffic patterns and based on the changes in their own available bandwidth. Strengthening bandwidth isolation, as mentioned above, is one approach to address this. Another effective approach is end-to-end encryption of network flows. Increased security awareness and privacy concerns among end-users – as well as the attempts by service providers to protect from competitors the valuable ancillary information created by end-users during service consumption – have

led to an increasing proportion of network traffic being encrypted⁷. However, end-to-end encryption negatively affects the capability of network infrastructure providers to perform content optimization, deploy TCP optimization proxies, implement caching functions, as well as deploy network security monitoring [37].

The relevant stakeholders must identify solutions that would allow to deploy the optimizations needed for network infrastructure scalability that can be effectively applied for encrypted traffic.

⁷OpenWave Mobility Press Release: <http://owmobility.com/press-release/over-80-of-traffic-on-mobile-networks-will-be-encrypted-in-12-months>

6 Conclusions

In this technical report we have reviewed the expectations towards 5G and its security considerations. It is apparent that the use-cases and business-cases for 5G will introduce a novel set of requirements and will cover a wide range of devices, beyond smartphones. 5G must be seen as an infrastructure of heterogeneous technologies, extending the support of devices over actual mobile telecommunications, with enhancements made not only in the radio access technology and core networks.

The service-oriented approach in 5G will require an open network architecture and increased logical separation in the 5G infrastructure, with 5G Slices consisting of a number network functions and radio access technology settings that support a specific set of use cases or business cases, and will cover all parts of the infrastructure. The 5G Slices will be enabled by API calls to the 5G network. This approach will require an increased dependence on software to control the infrastructure, and the trustworthiness of such software becomes vital for the success of 5G.

In this report we have briefly reviewed historical development of the authentication and key agreement protocol, as an example of incremental evolution of security in mobile telecommunications. We have also discussed known attacks on LTE networks, which must be mitigated in 5G. Furthermore, new actors and business models will require different trust models for 5G: the existing models assume a friendly environment between operators and devices, which is not true in today's networks – as demonstrated by the recent leaks. Hence, the security architecture for 5G must be aligned with new trust models. Lastly, the documented exceptions made from the 3GPP security working group recommendations, such as those mentioned in Section 5.2, must be accompanied by the exception's design rationale, and an analysis of its complications and any threats it impose.

We describe a proposed architecture for authentication, authorization and accounting for 5G. Today's authentication methods – which is the foundation of AAA with secure identification of services, end-points and users – will not scale with the expected boom of IoT devices. Furthermore, existing authorization methods are insufficient to support resource-constrained devices. The new architecture will bring a distributed approach for authentication and convey new models for granular access control decisions.

In this technical review, we have motivated the fact that incremental changes will not suffice in 5G to support the new use-cases, and new actors that bring new trust models. The AAA protocols will need to be extended to support the expected billions of IoT devices that will depend on 5G for communication. We have also provided a brief insight into software-defined networking and its security considerations. Finally, considering many of the future requirements of 5G can not be imagined today, the infrastructure and the security must be flexible enough to adapt to future needs.

References

- [1] NGMN Alliance, “NGMN 5G White paper.” https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf. Accessed February, 2016.
- [2] 5G Public Private Partnership, “5G Vision.” <https://5g-ppp.eu/wp-content/uploads/2015/02/5G-Vision-Brochure-v1.pdf>. Accessed February, 2016.
- [3] Radio Access and Spectrum, “FP7 – Future Networks Cluster “White paper on 5G radio network architecture.” http://fp7-semafour.eu/media/cms_page_media/9/SEMAFOUR_2014_RAScluster%20White%20paper.pdf. Accessed February, 2016.
- [4] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516, 2012.
- [5] The Economist Intelligence Unit, “Power to the patient: How mobile technology is transforming healthcare.” <http://www.economistinsights.com/analysis/how-mobile-transforming-healthcare>. Accessed February, 2016.
- [6] METIS, Mobile, “Wireless Communications Enablers for the Twenty-Twenty Information Society,” D1. 2, Initial channel models based on measurements, 2013.
- [7] Ericsson AB, “5G – Key component of the networked society, RWS-150009.” http://www.3gpp.org/news-events/3gpp-news/1734-ran_5g. Accessed February, 2016.
- [8] Qualcomm, “5G View on technology & standardization, RWS-150012.” http://www.3gpp.org/news-events/3gpp-news/1734-ran_5g. Accessed February, 2016.
- [9] ZTE, “Considerations on 5G Key technologies & Standardization, RWS-15002.” http://www.3gpp.org/news-events/3gpp-news/1734-ran_5g. Accessed February, 2016.
- [10] D. Talbot, “One Simple Trick Could Disable a City’s 4G Phone Network.” <https://www.technologyreview.com/s/507381/one-simple-trick-could-disable-a-citys-4g-phone-network/>. Accessed February, 2016.
- [11] V. Jungnickel, K. Manolakis, S. Jaeckel, M. Lossow, P. Farkas, M. Schlosser, and V. Braun, “Backhaul requirements for inter-site cooperation in heterogeneous LTE-Advanced networks,” in *Communications*

- Workshops (ICC), 2013 IEEE International Conference on, pp. 905–910, IEEE, 2013.
- [12] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, “Fog computing and its role in the internet of things,” in Proceedings of the first edition of the MCC workshop on Mobile cloud computing, pp. 13–16, ACM, 2012.
- [13] L. Yang, R. Dantu, T. Anderson, and R. Gopal, “Forwarding and control element separation (ForCES) framework,” tech. rep., RFC 3746, April, 2004.
- [14] A. Greenberg, G. Hjalmtysson, D. A. Maltz, A. Myers, J. Rexford, G. Xie, H. Yan, J. Zhan, and H. Zhang, “A clean slate 4D approach to network control and management,” ACM SIGCOMM Computer Communication Review, vol. 35, no. 5, pp. 41–54, 2005.
- [15] S. Jain, A. Kumar, S. Mandal, J. Ong, L. Poutievski, A. Singh, S. Venkata, J. Wanderer, J. Zhou, M. Zhu, et al., “B4: Experience with a globally-deployed software defined WAN,” in ACM SIGCOMM Computer Communication Review, vol. 43, pp. 3–14, ACM, 2013.
- [16] M. Casado, N. Foster, and A. Guha, “Abstractions for software-defined networks,” Communications of the ACM, vol. 57, no. 10, pp. 86–95, 2014.
- [17] N. Gude, T. Koponen, J. Pettit, B. Pfaff, M. Casado, N. McKeown, and S. Shenker, “NOX: towards an operating system for networks,” ACM SIGCOMM Computer Communication Review, vol. 38, no. 3, pp. 105–110, 2008.
- [18] T. Koponen, M. Casado, N. Gude, J. Stribling, L. Poutievski, M. Zhu, R. Ramanathan, Y. Iwata, H. Inoue, T. Hama, et al., “Onix: A Distributed Control Platform for Large-scale Production Networks.,” in OSDI, vol. 10, pp. 1–6, 2010.
- [19] V. Niemi and K. Nyberg, UMTS security. John Wiley & Sons, 2003.
- [20] U. Meyer and S. Wetzel, “A man-in-the-middle attack on UMTS,” in Proceedings of the 3rd ACM workshop on Wireless security, pp. 90–97, ACM, 2004.
- [21] U. Jang, H. Lim, and H. Kim, “Privacy-enhancing security protocol in LTE initial attack,” Symmetry, vol. 6, no. 4, pp. 1011–1025, 2014.
- [22] C. Lai, H. Li, R. Lu, and X. S. Shen, “SE-AKA: A secure and efficient group authentication and key agreement protocol for LTE networks,” Computer Networks, vol. 57, no. 17, pp. 3492–3510, 2013.

- [23] D. Forsberg, G. Horn, W.-D. Moeller, and V. Niemi, LTE security. John Wiley & Sons, 2012.
- [24] J.-K. Tsay and S. F. Mjøl̄snes, “A vulnerability in the umts and LTE authentication and key agreement protocols,” in Computer Network Security, pp. 65–76, Springer, 2012.
- [25] A. Shaik, R. Borgaonkar, N. Asokan, V. Niemi, and J.-P. Seifert, “Practical attacks against privacy and availability in 4G/LTE mobile communication systems,” arXiv preprint arXiv:1510.07563, 2015.
- [26] D. F. Kune, J. Koelndorfer, N. Hopper, and Y. Kim, “Location leaks on the GSM Air Interface,” ISOC NDSS (Feb 2012), 2012.
- [27] P. Langlois, “BLTE Pwnage: Hacking HLR/HSS and MME Core Network Elements.” <http://conference.hitb.org/hitbsecconf2013ams/materials/D1T2%20-%20Philippe%20Langlois%20-%20Hacking%20HLR%20HSS%20and%20MME%20Core%20Network%20Elements.pdf>. Accessed February, 2016.
- [28] 5G-Ensure Consortium, “Horizon 2020, call H2020-ICT-2014-2, proposal number 671562, 5G-ENSURE.” <https://5g-ppp.eu/5g-ensure/>. Accessed February, 2016.
- [29] GM Research & development, “Towards 5G – An Automotive Perspective, RWS-150011.” http://www.3gpp.org/news-events/3gpp-news/1734-ran_5g. Accessed February, 2016.
- [30] Z. A. Qazi, C.-C. Tu, L. Chiang, R. Miao, V. Sekar, and M. Yu, “SIMPLE-fying middlebox policy enforcement using SDN,” in ACM SIGCOMM computer communication review, vol. 43, pp. 27–38, ACM, 2013.
- [31] N. Paladi and C. Gehrman, “Towards Secure Multi-tenant Virtualized Networks,” in Trustcom/BigDataSE/ISPA, 2015 IEEE, vol. 1, pp. 1180–1185, IEEE, 2015.
- [32] R. H. Deng, J. Zhou, and F. Bao, “Defending against redirect attacks in mobile IP,” in Proceedings of the 9th ACM conference on Computer and communications security, pp. 59–67, ACM, 2002.
- [33] N. Paladi, “Towards Secure SDN Policy Management,” in Cloud Security and Privacy by Design, 2015 CloudSPD, 2015. In press.
- [34] P. Porras, S. Shin, V. Yegneswaran, M. Fong, M. Tyson, and G. Gu, “A security enforcement kernel for OpenFlow networks,” in Proceedings of the first workshop on Hot topics in software defined networks, pp. 121–126, ACM, 2012.

- [35] A. Wang, M. Iyer, R. Dutta, G. N. Rouskas, and I. Baldine, “Network virtualization: Technologies, perspectives, and frontiers,” Lightwave Technology, Journal of, vol. 31, no. 4, pp. 523–537, 2013.
- [36] S. Tripathi, N. Droux, T. Srinivasan, K. Belgaied, and V. Iyer, “Crossbow: A vertically integrated QoS stack,” in Proceedings of the 1st ACM workshop on Research on enterprise networking, pp. 45–54, ACM, 2009.
- [37] T. Anderson, P. Bosch, and A. Duminuco, “Bandwidth Control and Regulation in Mobile Networks via SDN/NFV-Based Platforms.” https://www.iab.org/wp-content/IAB-uploads/2015/08/MaRNEW_1_paper_26.pdf. Accessed February, 2016.