

Security in UMTS - Integrity

Telenor R&D

Runar Langnes (editor)

Tom E. Aamodt

Trond Friisø

Geir Kjøien

Øyvind Eilertsen

Version 1.00: 05. February 2001

Preface

This report is a delivery from the research project *MGSU01 UMTS security* financed by Telenor R&D. The project was running during year 2000, and was a part of the focus area mobile services of Telenor R&D. Project leader was Runar Langnes, and the other participants were Trond Friisø, Geir M. Kjøen, Tom Erling Aamodt and Øyvind Eilertsen.

Other deliveries from this project are reports concerning *authentication and key agreement, confidentiality, and the KASUMI algorithm*.

Table of contents

1	Introduction	4
1.1	Scope of document	4
1.2	References	4
1.3	Terminology	5
1.3.1	Acronyms	5
1.3.2	Definitions	7
2	UMTS security architecture	9
3	Objectives and principles	10
4	Threats against integrity	12
4.1	Risk assessment	12
5	Requirements	13
6	Services and mechanisms for integrity protection	14
6.1.1	Protection against transmission errors	14
6.1.2	Location of integrity protection, confidentiality protection and error control in the protocol stack.	14
6.2	Generation of integrity keys	15
6.3	Distribution of integrity keys	15
6.4	Algorithm for integrity protection	15
6.4.1	Protection against replay	16
6.4.2	Initialisation of synchronisation for integrity protection	16
6.4.3	Security mode set-up procedure	16
6.4.4	Integrity key selection when both service domains are involved	17
6.4.5	Key handling during re-authentication	18
6.4.6	Integrity key lifetime	18
6.5	Exceptions to integrity protection	18
7	Discussion	20
7.1	Unprotected parts of the system	20
7.1.1	User data	20
7.1.2	Fixed network	20
7.1.3	UE	21
7.1.4	Unprotected signalling messages	22
7.1.5	Denial-of-service attack	22
7.2	Strength of algorithm	22
7.3	False base station attack	22
7.4	Weaknesses in UMTS integrity protection mechanisms	23
8	Conclusions	24

1 Introduction

1.1 Scope of document

This report is one in a series of reports describing security features in UMTS, based on Release 99. The basic security services covered in these reports are authentication, confidentiality and integrity. This report addresses the integrity protection security service. Threats, requirements and mechanisms related to integrity are treated. Information is mainly collected from various 3GPP documents, and the intention is to give a presentation and analysis of the current status of the specifications regarding integrity protection, rather than a detailed description. Our aim is to consider whether the requirements, services and mechanisms are adequate to provide integrity of the information that should be protected from manipulation. Focus in this report is on integrity protection of signalling data in the access network. Security issues related to the mobile stations, USIM and applications running on these entities are outside the scope of the document. Neither are integrity services and mechanisms in the core network described since they are not part of Release 1999.

Most figures in this document have been found in the specifications from 3GPP.

1.2 References

Note that all 3GPP specifications referred to in this document are based on Release 99. These specifications can be found on <http://www.3gpp.org>

- [1] 3G TS 21.101 3rd Generation mobile system Release 1999 Specifications (Release 1999)
- [2] 3G TS 21.905 Vocabulary
- [3] 3G TS 21.133 Security Threats and Requirements
- [4] 3G TS 33.102 Security Architecture
- [5] 3G TS 33.120 Security Principles and Objectives
- [6] CCITT. *Data communication networks: Open systems interconnection (OSI); Security, structure and applications. Security architecture for open systems interconnection for CCITT applications*. Geneva, 1991. (X.800) (Technically aligned with ISO 7498-2: 1989)
- [7] ITU-T. *Data networks and open system communications. Security. Information technology - Open systems interconnection - Security framework for open systems: Integrity framework*. Geneva, 1995. (X.815) (Identical to ISO/IEC 10181-6: 1996)
- [8] T. E. Aamodt, T. Friisø G. Kjøen and R. Langnes. *Security in UMTS—Authentication and Key Agreement* Kjeller, Telenor FoU N 67/2000.

- [9] 3GPP TS 23.057 Mobile Station Application Execution Environment (MExE) ; Functional description
- [10] 3GPP TS 03.48: "Digital cellular telecommunications system (Phase 2+); Security Mechanisms for the SIM application toolkit ".
- [11] Eilertsen, Ø *Security in UMTS—The KASUMI algorithm*. Kjeller, Telenor FoU N 96/2000
- [12] T. Friisø, T. E. Aamodt, Ø. Eilertsen, G. Kjøen and R. Langnes. *Security in UMTS—Confidentiality* Kjeller, Telenor FoU N 81/2000
- [13] ISO/IEC 9797-1: 1999. *Information technology—Security techniques—Message Authentication Codes—Part 1: Mechanisms using a block cipher*. Geneva 1999
- [14] ITU-T. *Data networks and open system communications. Security. Information technology - Open systems interconnection - Security framework for open systems: Overview*. Geneva, 1995. (X.810) (Identical to ISO/IEC 10181-1: 1996)

1.3 Terminology

Terminology is adapted from various 3GPP specifications [1], primarily from the vocabulary [2].

1.3.1 Acronyms

<i>2G</i>	Second Generation (mobile system)
<i>3G</i>	Third Generation (mobile system)
<i>3GPP</i>	Third Generation Partnership Project
<i>AAL2</i>	ATM Adaptation Layer 2
<i>AES</i>	Advanced Encryption Standard
<i>AKA</i>	Authentication and Key Agreement
<i>ATM</i>	Asynchronous Transmission Mode
<i>AuC</i>	Authentication Centre
<i>CN</i>	Core Network
<i>CK</i>	Confidentiality Key
<i>CRC</i>	Cyclic Redundancy Check

<i>CS</i>	Circuit Switched
<i>DC</i>	Data Confidentiality
<i>DI</i>	Data Integrity
<i>GPRS</i>	General Packet Radio Service
<i>GSM</i>	Global System for Mobile communications
<i>HE</i>	Home Environment
<i>HFN</i>	Hyperframe Number
<i>HLR</i>	Home Location Register
<i>IK</i>	Integrity Key
<i>IMEI</i>	International Mobile station Equipment Identity
<i>IMSI</i>	International Mobile Subscriber Identity
<i>K</i>	Secret key stored in USIM and HLR/AuC
<i>KSI</i>	
<i>L3</i>	Layer 3 according to the OSI reference model
<i>MAC</i>	<i>MAC</i> has two meanings depending on the context: i) Link layer context: MAC=Medium Access Control or ii) Security context: MAC=Message Authentication Code [13]
<i>MAC-I</i>	Message Authentication Code used for data Integrity of signalling messages
<i>MAP</i>	Mobile Application Part
<i>ME</i>	Mobile Equipment
<i>MS</i>	Mobile Station
<i>MSC</i>	Mobile Switching Centre
<i>NAS</i>	Non Access Stratum
<i>OSI</i>	Open Systems Interconnection
<i>PHY</i>	Physical layer
<i>PS</i>	Packet Switched
<i>RANAP</i>	Radio Access Network Application Part
<i>RAND</i>	Random Challenge
<i>RLC</i>	Radio Link Control
<i>RNC</i>	Radio Network Controller

<i>RNSAP</i>	Radio Network Subsystem Application Part
<i>RRC</i>	Radio Resource Control
<i>SGSN</i>	Serving GPRS Support Node
<i>SN</i>	Serving Network
<i>SRNC</i>	Serving RNC
<i>UE</i>	User Equipment. A UE is an ME with a USIM
<i>UEA</i>	UMTS Encryption Algorithm
<i>UIA</i>	UMTS Integrity Algorithm
<i>UICC</i>	UMTS Integrated Circuit Card
<i>UMTS</i>	Universal Mobile Telecommunications System
<i>USIM¹</i>	UMTS Subscriber Identity Module
<i>UTRAN</i>	UMTS Terrestrial Radio Access Network
<i>VLR</i>	Visited Location Register

1.3.2 Definitions

The definitions presented in this report mainly comply with the definitions in X.800 [6], X.810 [14] and X.815 [7]. Some are extracted from [3].

control/signalling data

Signalling data comprises addresses, identities, location data, charging and billing data. Control data typically deals with resource management, call control and routing (corresponds to how the terms are used in [3])

cryptography

The discipline which embodies principles, means and methods for the transformation of data in order to hide its content, prevent its undetected modification and/or prevent its unauthorized use. [6]

data integrity

The property that data has not been altered or destroyed in an unauthorized manner. [6]

masquerade

The pretence by an entity to be a different entity. [6]

Message Authentication Code (MAC) [13]

¹ In some 3GPP specifications the definitions *Universal Subscriber Identity Module* and *User Service Identity Module* occur

An algorithm for computing a function which maps strings of bits and a secret key to fixed-length strings of bits, satisfying the following two properties:

- for any key and any input string the function can be computed efficiently;
- for any fixed key, and given no prior knowledge of the key, it is computationally infeasible to compute the function value on any new input string, even given knowledge of the set of input strings and corresponding function values, where the value of the i th input string may have been chosen after observing the value of the first $i-1$ function values.

private key

A key used with an asymmetric cryptographic algorithm and whose possession is restricted. (X.810)

user data

This type of traffic comprises all data transmitted on the end-to-end traffic channel by users to other users. The data could be digital data, voice, or any other kind of data generated by the user. (corresponds to how the term is used in [3])

2 UMTS security architecture

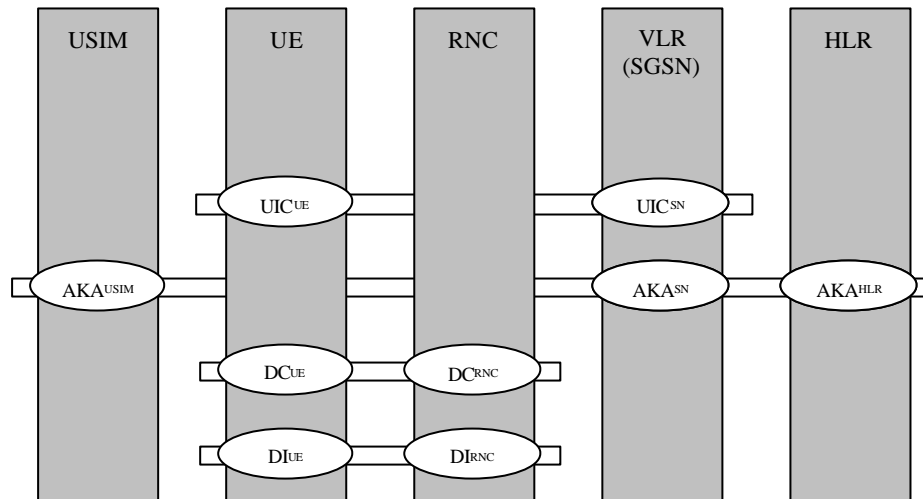


Figure 1: UMTS functional security architecture.

User Identity Confidentiality (UIC in Figure 1) is provided between the User Equipment (UE) and the Serving Network (SN) by using temporary identities. An enhanced version of this mechanism has been discussed, providing user identity confidentiality between the UE and the Home Environment (HE), but 3GPP has decided not to take this functionality into the specifications.

AKA – Authentication and Key Agreement – is the mechanism for mutual authentication and key agreement, including the functionality to trigger a re-authentication by the user, i.e. to control the access key pair lifetime. The access key pair includes the cipher and integrity keys, which are used to provide confidentiality and integrity between the UE and the RNC. AKA works between the USIM and the HLR/AuC. In practice, the SN is the counterpart to the USIM on behalf of the HLR/AuC. In some cases, the SN has enough information to perform the authentication without involving the HLR/AuC. These cases include authentication based on knowledge of a previously derived cipher/integrity key pair, and authentication by using an authentication vector, which has previously been transferred from the HLR/AuC to the VLR/SGSN in the SN. In addition to providing mutual authentication between the user and the network, the AKA procedure establishes an access key pair providing confidentiality and integrity in the access network.

Data Confidentiality (DC) is provided between the UE and the RNC for both *user* and *signalling data*. The cipher key used is derived during the AKA procedure.

Data Integrity (DI) is provided between the UE and the RNC for *signalling data*, but not for user data. Also the integrity key used is derived during the AKA procedure. The data integrity security service in UMTS is the subject for this report.

3 Objectives and principles

The main objectives and principles of UMTS security are specified in TS 33.120 [5]. High level objectives are:

- a) to ensure that information generated by or relating to a user is adequately protected against misuse or misappropriation;
- b) to ensure that the resources and services provided by serving networks and home environments are adequately protected against misuse or misappropriation;
- c) to ensure that the security features are adequately standardised to ensure world-wide interoperability and roaming between different serving networks and to ensure that the security features standardised are compatible with world-wide availability;
- d) to ensure that the level of protection provided to users and service providers is better than that provided in contemporary fixed and mobile networks;

Key principles are in prioritised order:

1. *Security elements within GSM and other second-generation systems that have proved to be needed and robust shall be adopted for UMTS.*
2. *UMTS will address and correct weaknesses in GSM.*
3. *UMTS will offer new security features (e.g. data integrity) and secure new services.*

The application of these principles to integrity is further discussed below.

In GSM (which is the relevant 2G reference system used throughout this paper) no integrity services is offered. Hence there is no security service to be retained according to principle 1. One 2G security feature relevant also for integrity is the principle that the operation of security features shall be independent of the user, i.e. the user does not have to do anything for the security features to be in operation.

Some security weaknesses have been detected in 2G systems. Among these are

- Active attacks using a false base station are possible
- Security data (e.g. CKs) are transmitted in clear between and within networks
- Data integrity is not provided except traditional non-cryptographic link-layer checksums
- Compromised authentication data (“triplets”) can be reused indefinitely. In UMTS these data² also comprise integrity keys so the concern applies to integrity as well
- Fraud and lawful interception were not considered in the design phase of 2G systems

Some security features are improved in 3G³ systems. These include:

- Key freshness assurance to the user at key agreement

² Five data elements in UMTS, hence the term “quintets” is used

³ In this report UMTS is the only 3G system treated

- Data integrity of signalling data between UE and RNC
- Integrity key length of 128 bits used for signalling data

Challenges that security services will have to cope with in 3G systems will probably be:

- Totally new services are likely to be introduced
- There will be new and different providers of services
- Mobile systems will be positioned as preferable to fixed line systems for users
- Users will typically have more control over their service profiles
- Non-voice services will be as important as, or more important than, voice services
- The terminal will be used as a platform for e-commerce and other sensitive applications

In chapter 5 we will identify requirements based on these objectives and principles in order to handle the threats introduced in the next chapter.

4 Threats against integrity

Threats against integrity are described in the specification TS 21.133 [3]. There is one generic threat against integrity, namely *manipulation of messages*. Manipulation includes deliberate or accidental modification, insertion, replaying, or deletion by an intruder. Data can be changed by accident or deliberately. Accidental change in data, like transmission errors, can be detected (and in some cases corrected) by non-cryptographic integrity mechanisms like Cyclic Redundancy Codes (CRC, see paragraph 6.1.1). Deliberate manipulation of data must be withstood by some kind of cryptographic mechanisms.

Both user data and signalling/control data are vulnerable to manipulation. In UMTS attacks may be conducted on the radio interface, in the fixed network or on the terminal and the USIM/UICC. The threats identified in [3] can be grouped and summarized to:

- **Manipulation of transmitted data:** Intruders may manipulate data transmitted (user traffic or signalling and control data) over the radio access interface, over other interfaces in the access network, over interfaces in the core network, or over the interface between the terminal and the UICC.
- **Manipulation of stored data:** Intruders may manipulate data that are stored on system entities, in the terminal or stored by the USIM. Access to the data may be obtained either locally or remotely, and may involve breaching physical or logical controls. The data vulnerable for manipulation also include the IMEI stored in the terminal, and data and applications downloaded to the terminal or USIM.
- **Manipulation by masquerading:** Intruders may masquerade as a communication participant and thereby manipulate data on any interface. Another attack is to manipulate the USIM behaviour by masquerading as the originator of malicious applications or data downloaded to the terminal or USIM.

4.1 Risk assessment

In risk analysis one is interested in the combination of severity of impact and likelihood of occurrence of an event. [3] presents some threats evaluated as being of major or medium value in risk analysis carried out in 3GPP. Extensive use has been made of the collected experiences of operators of first generation (analogue) systems and second generation systems (especially GSM) as regards current and envisaged threats to mobile systems. Threats of relevance for integrity are summarised below. Some of them are indicated to be of major severity.

- **Manipulation by masquerading:** On the radio interface this is considered to be a major threat, whereas manipulation of the terminal or USIM behavior by masquerading as the originator of applications and/or data is considered to be of medium significance.
Masquerading could be done both to fake a legal user and to fake a serving network.
- **Manipulation of stored data:** Only the risks associated with the threats to data stored on the terminal or USIM/UICC are regarded to be significant, and only the risk for manipulation of the IMEI is regarded as being of major importance.

5 Requirements

Overall security objectives for 3G systems are described in [5] and the requirements are outlined in [3]. It is a main objective that 3G security shall retain security elements from 2G systems (GSM) where these are considered sufficient. Security elements not considered sufficient in 2G systems shall be improved or replaced by better security features and mechanisms.

Requirements on system integrity are according to [3]:

- **It shall be possible to protect against unauthorised modification of transmitted data** (user traffic, signalling and control data). The radio access link is considered to be most exposed for attacks and therefore the requirement particularly apply to this interface.

It is assumed that user traffic contains sufficient redundancy such that a stream cipher provides a basic level of data integrity protection on the radio interfaces. If that is not sufficient and additional measures are required, the application should be aware and measures should be implemented at the application layer.

- **It shall be possible to protect against unauthorised modification of user-related data.** This can be data downloaded to or stored in the terminal or in the USIM or data stored or processed by a provider.
- **It shall be possible to ensure that the origin and integrity of applications and/or data downloaded to the terminal and/or the UICC can be checked.** It may also be necessary to ensure the confidentiality of downloaded applications and/or data.
- **It shall be possible to ensure the origin, integrity and freshness of authentication data,** particularly of the cipher key on the radio interface.
- **It shall be possible to secure infrastructure between operators .**

6 Services and mechanisms for integrity protection

The UMTS integrity algorithm (UIA) shall be implemented in the UE and in the RNC. Integrity protection shall apply at the RRC layer. Only control data is protected against loss of integrity. The data integrity of logical channels for user data is not protected. Primarily two kinds of signalling messages are transported over the radio interface: RRC-generated signalling messages and NAS messages generated in the higher layers. The latter category of messages is transported transparently through UTRAN. Examples include Mobility Management (MM) and Call Control (CC) messages. After the RRC connection has been established and the security mode set-up procedure has been performed, all dedicated control signalling messages between UE and the network shall be integrity-protected.

Throughout this document we use the notion “integrity protection” when protection is applied to the level of RRC messages. But first we introduce a related mechanism on lower layers.

6.1.1 Protection against transmission errors

Violation to integrity can be caused by deliberate attack and by accidental distortion. On the radio link signals can be severely exposed to noise, and *error control mechanisms* are introduced to reduce the effect of this phenomenon. These mechanisms are usually implemented in the link layer protocols or even on the physical layer. In the case of ATM/AAL5 (used in the bearer of signalling protocols in the fixed part of UTRAN) a 32-bit Cyclic Redundancy Code (CRC) is computed for each 48-byte cell. On the UTRAN radio interface redundancy of varying size is added to transport blocks on the physical layer, and the result of the error control is handed up to layer 2 where appropriate actions are taken.

6.1.2 Location of integrity protection, confidentiality protection and error control in the protocol stack.

Integrity protection shall be applied on a per-message basis. Therefore it has to take place at the Radio Resource Control (RRC) layer. Encryption will be applied in the Medium Access Control (MAC) sublayer or in the Radio Link Control (RLC) sublayer of the data link layer. The actual sublayer is depending on RLC mode. Error control is performed on the physical layer (PHY) in cooperation with the RLC sublayer.

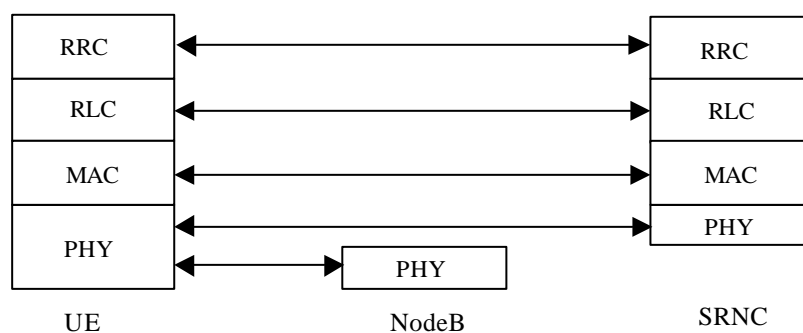


Figure 2: Protocol layers between UE and RNC

6.2 Generation of integrity keys

Generation and distribution of keys for integrity and confidentiality are part of the AKA procedure described in [8]. The 128-bit integrity key, IK, is not distributed to the mobile station, but is generated independently in the AuC and in the USIM. The integrity key generating function, which in the UMTS specifications is denoted “f4”, takes a random challenge RAND and the long-term secret key K as input parameters. K is shared between the USIM and the AuC and it is fundamental to all the security services in UMTS that this key is not unveiled. RAND is generated in the AuC and received by the USIM during the authentication procedure.

6.3 Distribution of integrity keys

The integrity protection in UMTS is implemented between the RNC and the UE. Therefore IK must be distributed from the AuC to the RNC. The IK is part of an *authentication vector* which is sent to the SN (VLR/SGSN) from the AuC following an *authentication data request*. To facilitate subsequent authentications, up to 5 authentication vectors are sent for each request. The IK is sent from the VLR/SGSN to the RNC as part of a RANAP message called *security mode command*.

To enable the communication to proceed at handover, the IK is transmitted within the network infrastructure from the old RNC to the new RNC. The IK remains unchanged at handover. Hence, the IK will in some situations be transmitted in clear over the Iur interface. If a user has connections to both circuit-switched and packet-switched domains he will have been authenticated independently in each domain. As stated earlier, distribution of integrity and confidentiality keys is part of the AKA procedure [8]; therefore, separate sets of integrity keys are maintained for CS and PS domains.

6.4 Algorithm for integrity protection

The UMTS integrity algorithm f9 takes as input the integrity key IK, the MESSAGE that is to be protected, and three additional parameters:

COUNT-I a time variant parameter (32 bits);
DIRECTION direction of transmission (1 bit);
FRESH a random number generated in the RNC (32 bits).

Based on these inputs the sending side computes a 32-bit Message Authentication Code MAC-I and sends it along with the message. The receiving side computes an expected value XMAC-I from the same inputs and verifies the integrity of the MESSAGE by comparing XMAC-I with the received MAC-I.

A 4-bit algorithm identifier is defined to support different integrity algorithms. Currently only the KASUMI-based integrity algorithm (UIA1) has been assigned a value.

6.4.1 Protection against replay

The input parameter COUNT-I protects against replay during a connection. COUNT-I consists of two parts: The 28 most significant bits are the HYPERFRAME NUMBER (HFN), while the 4 least significant bits constitute the RRC Sequence Number. The largest HFN from the previous connection is stored in the USIM. During connection set-up this value is incremented by one and sent from the UE to the network. In this way the user can be assured that no COUNT-I value is re-used (by the network) with the same integrity key.

An integrity key may be used for several consecutive connections. The FRESH parameter assures the network side that the user is not replaying old MAC-Is. At connection set-up the RNC generates a random value FRESH and sends it to the user in the (RRC) *security mode command* (point 7 in Figure 3). This value of FRESH is subsequently used by both the network and the user throughout the duration of a single connection.

To sum up:

- The value of COUNT-I is incremented for each message, while the generation of a new FRESH value and *initialization* of COUNT-I take place at connection set-up.
- The COUNT-I value is initialized in the UE and therefore primarily protects the user side from replay attacks. Likewise the FRESH value primarily provides replay protection for the network side.

6.4.2 Initialisation of synchronisation for integrity protection

The ME and the USIM contain starting values for the HFN part of the COUNT-I. These are called $START_{CS}$ and $START_{PS}$ for CS and PS domains respectively. The ME only contains valid START values when it is powered on and the USIM is inserted. After power-on or insertion of a USIM, the USIM sends its START values to the ME, and the ME stores them. During idle mode, the START values in the ME and in the USIM are identical and static. When the ME is powered-off or the USIM is removed, the ME deletes its START values.

6.4.3 Security mode set-up procedure

The VLR/SGSN initiates integrity protection (and encryption) by sending the RANAP message *security mode command* to the SRNC. This message contains a list of allowed integrity algorithms and the IK to be used. Figure 3 shows how the input parameters for the integrity algorithm are interchanged between the communicating nodes. (Since the keys and algorithm identifiers for integrity protection and confidentiality are distributed in the same messages both parts are shown in the figure.) Since the UE can have two ciphering and integrity key sets (for the PS and CS domains, respectively), the network includes a Core Network type indicator in the *security mode command* message.

The *security mode command* to UE (point 7 in Figure 3) starts the downlink integrity protection, i.e. all subsequent downlink messages sent to the UE are integrity protected. The *security mode complete* from UE (point 9 in Figure 3) starts the uplink integrity protection, i.e. all subsequent messages sent from the UE are integrity protected.

The network must have the “UE security capability” information before the integrity protection can start, i.e. the “UE security capability” must be sent to the network in an

unprotected message. Returning the “UE security capability” to the UE in a protected message later on (step 7 in Figure 3) will allow UE to verify that it was the correct "UE security capability" that reached the network.

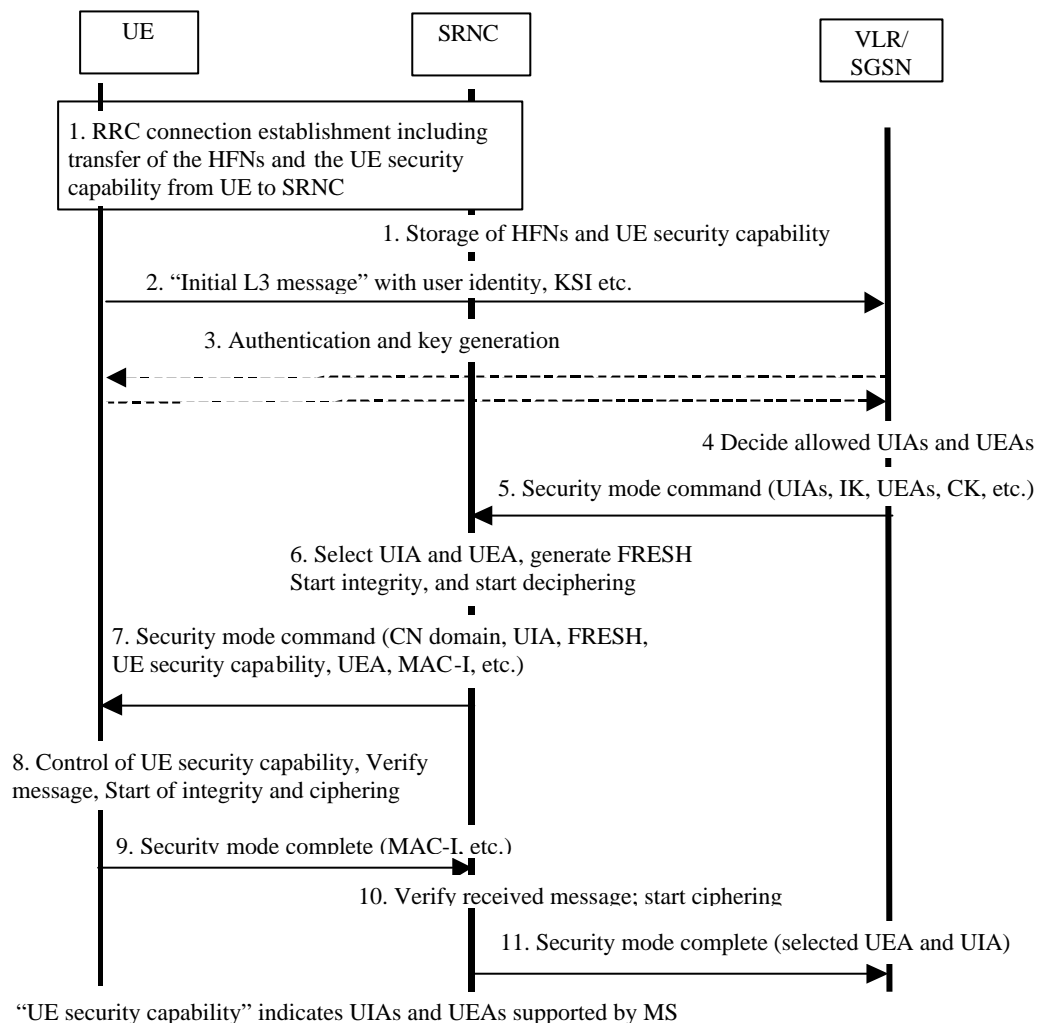


Figure 3: Security mode set-up (source TS.33.102)

The procedure outlined above describes a straightforward and successful case of security set-up. If the UE and the network have no versions of the integrity algorithm in common, the connection shall be released.

The supervision of failed integrity checks shall be performed both in the UE and the SRNC. If a failed integrity check (i.e. faulty or missing MAC-I) is detected after the integrity protection is started the concerned message shall be discarded. This can happen on both the RNC side and the UE side.

6.4.4 Integrity key selection when both service domains are involved

As stated earlier, separate integrity keys for the PS and CS domains are stored in USIM and SRNC. Signalling data for services delivered by either of the service domains is sent over common logical (signalling) channels. These logical channels are data integrity-protected using the IK of the service domain for which the *most recent* security mode negotiation took place.

Thus the integrity key of an ongoing (already integrity-protected) signalling connection must be changed when a new RRC connection (with another service domain) is established.

6.4.5 Key handling during re-authentication

A change in integrity key may also be required when a security mode negotiation follows a re-authentication during an ongoing connection. Re-authentication may be performed implicitly, in which case knowledge of the correct keys is taken as proof of identity. Re-authentication may also be performed based on the next authentication vector stored in the SN. In this case a new IK will be taken into use.

6.4.6 Integrity key lifetime

To avoid attacks using compromised keys, a mechanism is needed to ensure that a particular integrity key⁴ set is not used for an unlimited period of time. Each time an RRC connection is released, the values $START_{CS}$ and $START_{PS}$ of the bearers that were protected in that RRC connection are stored in the USIM. When the next RRC connection is established these values are read from the USIM.

The operator shall decide on a maximum value for $START_{CS}$ and $START_{PS}$. This value is stored in the USIM. When the maximum value has been reached, the cipher key and integrity key stored on USIM shall be deleted, and the ME shall trigger the generation of a new access link key set (a cipher key and an integrity key) at the next RRC connection request message.

6.5 Exceptions to integrity protection

There are three exceptions to the mandatory start of integrity protection:

- If the only purpose with the signalling connection establishment and the only result is periodic location registration, i.e. no change of any registration information.
- If there is no UE-MSC/VLR (or UE-SGSN) signalling after the initial L3 signalling message sent from UE to MSC/VLR (or SGSN), i.e. in the case of deactivation indication sent from the UE followed by connection release.
- If the only UE-MSC/VLR (or UE-SGSN) signalling after the initial L3 signalling message sent from UE to MSC/VLR (or SGSN), and possible user identity request and authentication (see below), is a reject signalling message followed by a connection release.

When the integrity protection shall be started, the only procedures between UE and MSC/VLR respective SGSN that are allowed after the initial connection request (i.e. the initial Layer 3 message sent to VLR/SGSN) and before the security mode set-up procedure are the following:

- Identification by a permanent identity (i.e. request for IMSI)

⁴ The same applies to confidentiality keys

- Authentication and key agreement

After completion of security mode set-up procedure, all signalling messages except the following shall be integrity protected:

- Paging Type 1
- RRC Connection Request
- RRC Connection Set-up
- RRC Connection Set-up Complete
- RRC Connection Reject
- System Information (broadcasted information)
- Handover to UTRAN complete.

7 Discussion

In this report we have presented objectives, requirements and threats identified by the 3GPP specification group and documented in the respective 3GPP specifications. We have decided not to challenge these statements and limit ourselves to discuss whether services and mechanisms in UMTS meet the integrity related objectives and requirements deduced from the identified threats.

7.1 Unprotected parts of the system

Cryptographic integrity protection in UMTS is implemented in only a limited part of the system. Only signalling data are protected, and the protection is limited to the transmission between the SRNC and the UE. In addition there are some cases where even signalling messages over the air interface are left unprotected.

7.1.1 User data

The fact that user data messages have no associated integrity checksum does not mean that user data is totally unprotected against manipulation. Firstly, the error control mechanisms (see 6.1.1) are able to detect and possibly correct some changes in transmitted data. Secondly, and more important: It is fair to assume that user data will have sufficiently low entropy for manipulation to be discovered by the fact that the output from the decryption process most likely should give no meaningful information. In the last case, however, it will be up to the user or an application to detect and take proper action upon reception of manipulated data.

The assumption above might be valid for speech and written prose. For user data with high uncertainty of an outcome, e.g. figures in a bank transaction, the low-grade integrity protection implicitly offered by encryption/decryption is not satisfactory. In that case an end-to-end integrity protection service would be the most appropriate remedy.

7.1.2 Fixed network

The protocols for distributing integrity keys are:

- MAP (between HLR/AuC and the VLR/SGSN in the core network)
- RANAP (over the Iu-interface between VLR/SGSN and serving-RNC)
- RNSAP (over the Iur-interface between RNCs at some handover situations)

In Release99 these keys will be transported unprotected both in the core network and over the Iu/Iur-interfaces. This represents a vulnerability that operators should consider carefully. Transmission lines could be protected by physical means or by point-to-point encryption.

In earlier stage of the security specification work it was assumed that the VLR/SGSN was the end point for integrity and confidentiality protection. Since this is no longer the case, security-related data must be protected on the Iu interface. The challenge of securing the core network (where AuC and VLR/SGSN are sited) was of course recognized, but the Iu interface is equally vulnerable to attacks. If RNCs are designed to handle large networks and possibly co-

located with VLR/SGSN, protection of signalling data could be relatively easy to achieve. But in a network with a great number of distributed RNCs for every VLR/SGSN, securing the Iu interface is a more challenging task.

7.1.3 UE

Some of the threats against integrity of data in the mobile terminal and USIM/UICC are judged to be of major severity. In this study we have not gone deeply into these threats. We have focused on the basic security services in the network access domain (marked with I in Figure 4). Some of the security issues related to UE belong to what is termed as user (III) and application (IV) domains. Figure 4 indicates these security domains.

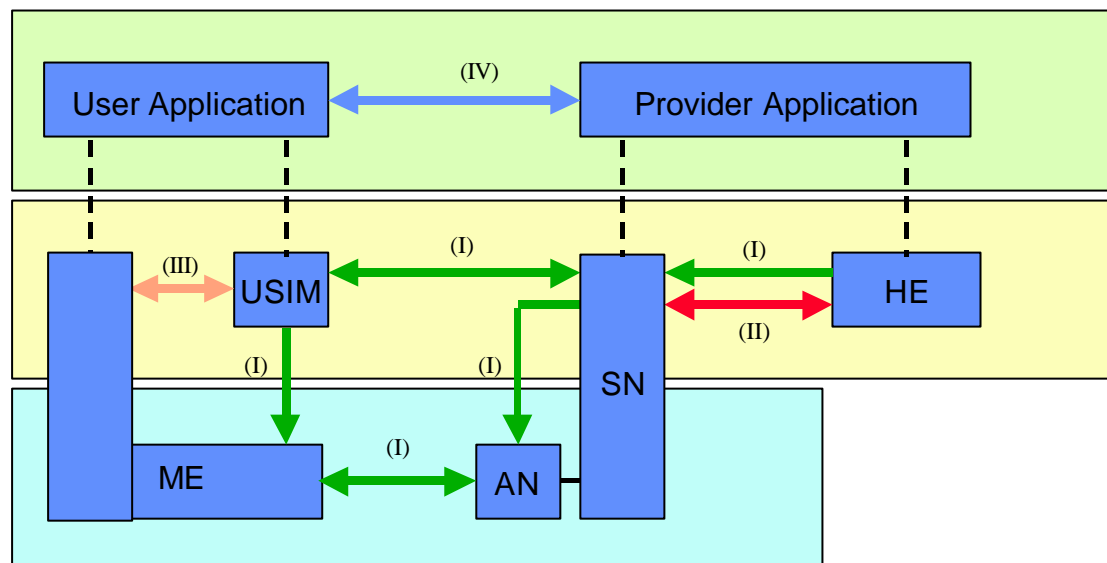


Figure 4: Overview of the security architecture (source [4])

In this report we restrict ourselves just to mention specifications relevant for the user and application domains. Two “platforms” or execution environments for applications are specified within 3GPP. User Equipment (UE) includes both the equipment from the vendor of the mobile terminal, called ME, and the smart card, UICC, on which USIM is the most essential application. For USIM the USIM Application Toolkit, USAT, is specified while the Mobile station application Execution Environment, MExE, specifications apply for the mobile terminal, ME. Within MExE [9] four⁵ different security levels or *domains* are defined (not to be confused with the domains just introduced). Applications signed by the network operator are permitted to run in the most highly trusted domain. In the manufacturer domain only applications signed by the manufacturer of the ME have permission to run. In the third party domain only applications endorsed by trusted third parties can execute. Untrusted MExE executables are not in a specific domain, and have very reduced privileges. This class of applications has restricted access authorisation on the ME or USIM, and the intention is that they shall not have certification permitting access to MExE UE’s capabilities. WAP applications belong to this class because WAP applications have no signing functionality

⁵ To be strict just three levels are called “domains” and the fourth level is considered outside the domains

specified. Of course, from a security point of view one might fear that the popularity of such untrusted applications will push the limits for what privileges will be offered in this untrusted domain.

USAT provides a standardised execution environment for applications stored on the USIM/UICC and the ability to utilize certain functions of the supporting mobile equipment. A transport mechanism is provided enabling applications to be downloaded and/or updated. For details about how security is taken care of in USAT see [10].

7.1.4 Unprotected signalling messages

Identification by a permanent identity and AKA may take place prior to the security mode set-up procedure. The unprotected transmission of permanent identity is primarily a threat against user identity confidentiality (see [12]). But it is also a possibility that the user identity could be tampered with. This is a far more complicated intervention than just eavesdropping, and it is not obvious how this could be exploited beyond creating general disturbance in the system.

During the key agreement phase a possibility for denial-of-service attack occurs (see below).

7.1.5 Denial-of-service attack

In the security mode set-up procedure the security capabilities of the mobile station are transmitted unprotected to the RNC. This may open for attacks. Assume for instance that an attacker change the security capabilities in the first message from the UE to the RNC. The following verification of the integrity of the UEs service capabilities stored in the RNC will fail, and the connection procedure will terminate. Hence, this is a possible denial-of-service attack.

7.2 Strength of algorithm

The integrity function employs the KASUMI [11] algorithm with a 128-bit key. The key-length makes the algorithm resistant to all known attacks based on exhaustive keysearch. It is however necessary to use a good random number generator for key generation, but this should be achievable for any serious operator. (Nevertheless, for several GSM operators the key generation algorithm has turned out to be a weak point.) A key generation algorithm built on the newly chosen AES algorithm (MILENAGE) is recommended by 3GPP.

7.3 False base station attack

This kind of attack is mainly prevented by the mutual authentication procedure. One threat related to false base stations is the threat of channel hijacking. This could be carried out by an intruder operating as man-in-the-middle between a target user and a false base station during call setup. After authentication encryption could have been disabled and the intruder could take over the call. Since encryption can be disabled (in some countries the authorities will probably not allow encryption), channel hijacking could have been a serious threat in absence of integrity protection. However, integrity protection of signalling data is mandatory in UMTS, and constitutes a second line of defence towards false base station attacks.

7.4 Weaknesses in UMTS integrity protection mechanisms

To sum up, the main weaknesses in UMTS integrity protection are:

- Integrity keys used between UE and RNC generated in VLR/SGSN are transmitted *unencrypted* to the RNC (and sometimes between RNCs)
- Integrity of user data is not offered
- For a short time during signalling procedures, signalling data are unprotected and hence exposed to tampering.

8 Conclusions

One of the weaknesses pointed out in second generation mobile systems is the lack of integrity protection. The introduction of an integrity protection service in UMTS represents a substantial improvement compared to GSM. Integrity is provided to signalling data between the RNC and the UE. In some cases it is expected that ciphering is not used. This can for instance be due to legal restrictions in some countries. In absence of encryption, integrity protection of signalling data provides protection against channel hijacking.

User data is not specifically protected against manipulation. A basic integrity protection is inherently provided by the confidentiality service. This might be sufficient for some kinds of user data traffic. However, user data sensitive to manipulation should be protected by an end-to-end integrity protection mechanism.

Core network traffic is neither integrity nor confidentiality protected. Hence, signalling data and even keys for integrity protection is transported in clear. Core network security is addressed in ongoing 3GPP specification work for subsequent releases of UMTS.

Integrity for data on terminals and on USIM/UICC is only very briefly discussed in this document. We have primarily focused on the network access domain. As offering of services and applications in UMTS gets more and more attention the possibility of hostile applications showing up is increasing. Therefore integrity protection in the application domain will be important.