Security Mechanisms in UMTS

Stefan Pütz, Roland Schmitz, Tobias Martin

This contribution presents an overview of the security of the 3rd generation mobile radio system UMTS as currently standardised by the 3rd Generation Partnership Project 3GPP. We discuss the underlying principles and show to which extent the security of 2nd generation systems as GSM is improved and enhanced by UMTS. The UMTS Authentication and Key Agreement protocol, the security algorithms deployed for UMTS and the interworking mechanisms between 2nd and 3rd generation systems are described in detail.[1]

Dr. Stefan Pütz
Promotion in IT Security, Leiter IT Security Networks bei T-Mobil, Vice-chairman der 3GPP Security Group. Arbeitsgebiete: IT Security, insb. UMTS Security, Standardisierung.
E-Mail: stefan.puetz@t-mobil.de

Dr. Roland Schmitz
Promotion in Mathematik, seit 1995 bei T-Nova, Arbeitsgebiete: Sicherheit mobiler Kommunikation, Standardisierung digitaler Signaturen.
E-Mail: roland.schmitz@t-systems.de

Dipl. Math. Tobias Martin

Mitarbeiter im Bereich Informationssicherheit von T-Nova, Mitglied von ETSI SAGE
E-Mail: tobias.martin@t-systems.de

[1] This contribution is based on [VIS2001].

1 Introduction

During 1998 a worldwide harmonisation and globalisation process for the 3G (3rd generation) mobile radio systems has taken place. Therefore, the standardisation process for UMTS (Universal Mobile Telecommunications System) has moved from ETSI (European Telecommunications Standards Institute) to 3GPP (3rd Generation Partnership Project). 3GPP was set up by various regional standards organisations and other related bodies from different continents and countries, e.g. Europe (ETSI), Japan (ARIB, TTC), Korea (TTA), China (CWTS) and North America (T1), that have agreed to co-operate for the production of a complete set of globally applicable technical specifications for a 3G mobile radio system based on evolution of the GSM (Global System for Mobile Communications) core network.

3G mobile radio systems will involve more players, e. g. content and service providers, and more operators, which will result in complex system and roaming scenarios. 3G systems will exist of and interact with a lot of different network types. Also 3GPP will promote wireless as preferred means of communication. To ensure that UMTS will work securely and reliably under this condition, the 3GPP security group was established in early 1999 to define the UMTS security. The standardisation of UMTS security within 3GPP has now reached a reasonably stable state.[2] It is attempted in this paper to provide a fairly complete overview of the UMTS security by highlighting its new security features and describing the most essential mechanisms (incl. protocols and algorithms) deployed to provide these features.

[2] 3GPP focused on the first release to be frozen end of year 1999 (called Release 99). During the year 2000 a lot of corrections and functional modifications became necessary.

2 3G Security Principles

The scope of 3G security is formed by a few principles. These principles state what is to be provided by 3G security as compared to the security of 2G (2nd generation) systems [TS 33.120].

- Build on the security of 2G systems
  Security elements within GSM and other 2G systems that have proved to be needed and robust shall be adopted. Furthermore compatibility with GSM in order to ease interworking and handover shall be ensured.
- Improve on the security of 2G systems
  3G security will address and correct real and perceived weaknesses in GSM and other 2G systems.
- Offer new security features
  3G security will secure new services offered by 3G systems and take account of changes in network architecture.

2.1 Building on GSM Security

The well known security features of GSM are described in the ETSI standards [GSM 02.09, GSM 03.20]. Furthermore, the GSM security architecture and mechanisms are explained in [DuD1996]. In the ten years of being in operation, the security architecture of GSM has proven to provide the GSM subscribers a sufficient level of security. The big commercial success of GSM is at least partly owed to its robust security architecture, which has struck a good balance between subscriber's needs and commercial interests. It was therefore a quite natural decision for the 3GPP security group, to retain basic security features of GSM for 3G, namely:

- Subscriber identity confidentiality
- Subscriber to network authentication
- Radio interface encryption
- Use of a SIM (Subscriber Identity Module) card as a terminal independent secu-

rity module, consisting of removable hardware

- Authentication of subscriber towards the SIM
- Security operation without user assistance
- An authentication procedure, that is performed by the Serving Network (SN), but nevertheless requires minimal trust in the SN, because no permanent user-individual keys are transferred from the user's Home Environment (HE) to the SN.
- The possibility to use operator-individual authentication algorithms

3 Additional UMTS Security Features

In addition to the well-known GSM security features given in 2.1, UMTS provides the following security features [DuD 2001]:

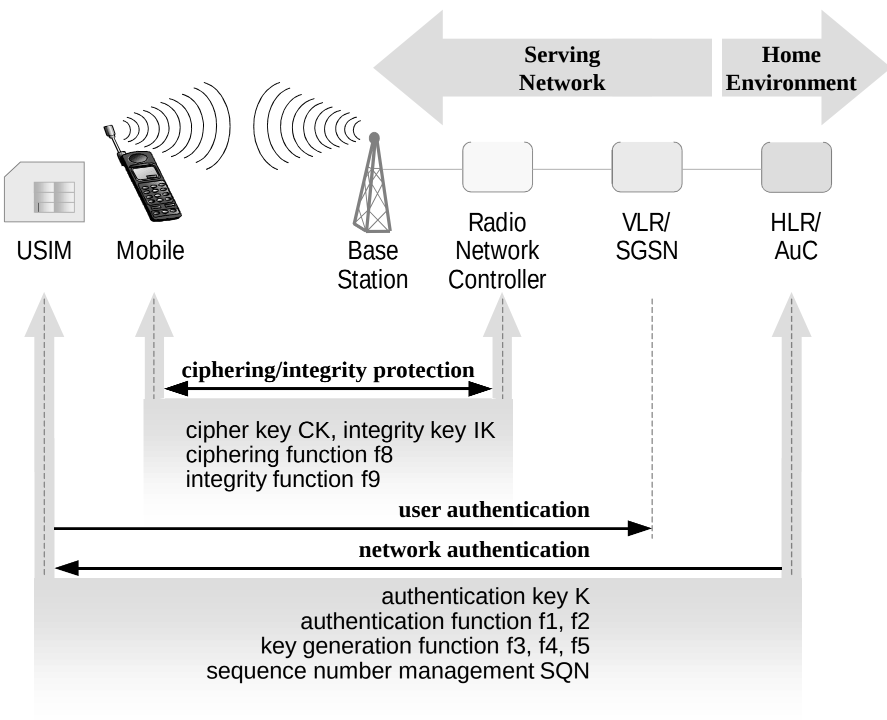

Fig. 1: UMTS Security Architecture

Enhanced UMTS Authentication and Key Agreement Mechanism

The UMTS authentication and key agreement mechanism provides the following features on top of the corresponding GSM mechanism:

- HE to USIM (UMTS Subscriber Identity Module)[3] authentication (cf. 4.3)
- Sequence number management in order to prevent resp. ensure controlled re-use of authentication vectors (cf. 4.4)
- AMF (Authenticated Management Field) providing a secured channel from HE to USIM for defining operator-specific options in the authentication process (cf. 4.2)
- Agreement on an Integrity Key (IK) used for integrity protection of signalling information (cf. 4.2)

Integrity Protection of Signalling Information

Integrity protection of signalling information serves to secure connection establishment and provides in-call (or local) authentication independent of ciphering. Thereby this is one mean to effectively prevent so called false-base-station-attacks.

USIM Control of Cipher/Integrity Key Usage

The USIM is enabled to keep track of the amount of data secured using a particular cipher/integrity key pair and triggers a new authentication at the next connection set-up when the amount of data already secured by the current key pair exceeds a certain threshold.

SN Control of Cipher/Integrity Key Lifetime

The SN triggers refreshment of cipher/integrity keys on a regular basis to control and limit the usage of cipher/integrity key pairs.

Ciphering/Integrity Protection

Ciphering/integrity protection terminates at the RNC (Radio Network Controller)[4], which ensures that the interface between base station and RNC is secured.

Cipher/integrity key lengths of 128 bits provide a security margin for future advances in computing power.

Trust and Confidence by Published Algorithms

Reviewed and public 3G security algorithms for ciphering, integrity protection and authentication will foster trust and confidence in the security of the UMTS system.

Authentication Failure Indication

In case of a failed authentication attempt, the reason is signalled back to the HE, which may help to detect intruders masquerading as legitimate networks.

4 Authentication and Key Agreement

The UMTS Authentication and Key Agreement (AKA) protocol has been designed in such a way that the compatibility with GSM is maximised and a migration from GSM to UMTS is easily possible (cf. section 6). The basic architecture of a symmetric challenge-response protocol, as it is deployed in GSM, has therefore been retained for UMTS. However, as already mentioned in section 3, there are significant enhancements to the related GSM protocol which serve to achieve additional protocol goals:

- Authentication of HE to the user
- Agreement on an Integrity Key (IK) between user and SN
- Mutual assurance of freshness of agreed Cipher Key (CK) and Integrity Key (IK) between SN and user

The UMTS AKA as designed by 3GPP has also been adopted by a competing, mostly

[3] USIM is in some sense the UMTS analogon to the GSM SIM card.

[4] RNC is the UMTS analogon to the GSM network entity called Base Station Controller.

north-american based standardisation effort called 3GPP2 in order to ensure the possibility of global roaming for 3G subscribers. In what follows, we will briefly describe how the UMTS AKA protocol works and how the protocol goals are achieved. [TR 33.902] provides a formal analysis of the UMTS AKA. The brief description given here is partly based on the detailed account of the UMTS AKA in [ISSE2000].

4.1 Message Flow

The UMTS AKA is based on the assumption that the Authentication Center (AuC) of the user's home environment and the user's USIM share a user specific secret key K, certain message authentication functions *f1, f2* and certain key generating functions *f3, f4, f5*. The UMTS AKA consists basically of two phases:

- Generation of Authentication Vectors
  After receiving an authentication data request from an SN, the HE/AuC generates an array of *n* authentication[5] vectors, each consisting of the following five components: A random number RAND, an expected response XRES, a cipher key CK, an integrity key IK and an authentication token AUTN. This array of *n* authentication vectors is then sent to the requesting SN.
- Authentication and Key Agreement
  In an authentication exchange the SN, resp. one of its corresponding network entities, namely Visitor's Location Register (VLR) or Serving GPRS Support Node (SGSN), selects the next (the *i*-th, where $1 \leq i \leq n$) authentication vector from the ordered array and sends RAND(*i*), AUTN(*i*) to the user. The USIM checks whether AUTN(*i*) can be accepted, i.e. whether AUTN(*i*) constitutes a valid authentication token, and if so, produces a response RES(*i*) which is sent back to the SN, which compares RES(*i*) to XRES(*i*). The USIM now also computes CK and IK which are subsequently used for ciphering and integrity protection on the air interface.

We will look at these two phases in greater detail in the following two subchapters.

4.2 Generation of Authentication Vectors

Upon receipt of an authentication data request, the HE/AuC starts with generating a fresh sequence number SQN and an unpredictable challenge RAND. For each user the HE/AuC keeps track of a counter: SQN_{HE}

Subsequently the following values are computed by the HE/AuC by using the user-specific key K and an operator-specific Authentication Management Field AMF[6] (cf. figure 2):

- A message authentication code MAC = $f1_K$(SQN || RAND || AMF) where f1 is a message authentication function
- An expected response XRES = $f2_K$(RAND) where f2 is a (possibly truncated) message authentication function
- A cipher key CK = $f3_K$(RAND) where f3 is a key generating function
- An integrity key IK = $f4_K$(RAND) where f4 is a key generating function
- An anonymity key AK = $f5_K$(RAND) where f5 is a key generating function

Finally the authentication token AUTN = SQN ⊕ AK || AMF || MAC is constructed.

Here, AK is an anonymity key used to conceal the sequence number as the latter may expose the identity and location of the user. The authentication token AUTN is sent to the user together with the random challenge RAND by the SN. Its purpose is twofold: Firstly, it authenticates the HE to the user, since AUTN can only be computed by an entity in possession of K. It cannot be replayed, because the time-variant parameter SQN is included in the computation of AUTN. Secondly, by verifying that AUTN is correct, the user is also assured that the serving network is trusted by the user's HE (cf. 4.3.2).

The parameters (RAND, XRES, CK, IK, AUTN) together form the UMTS authentication vector sent to the SN by the HE.

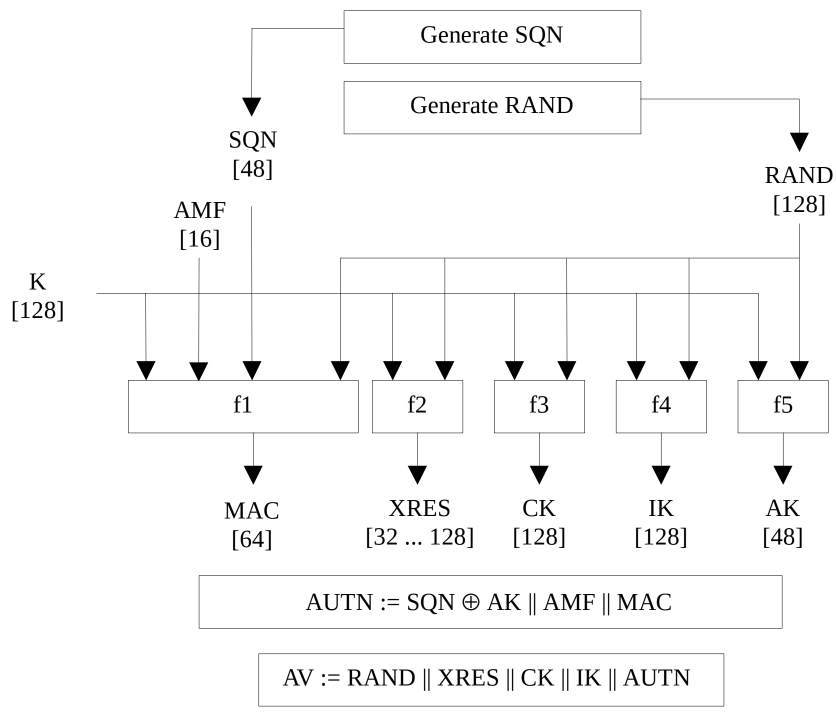

Fig. 2: Generation of authentication vectors [TS 33.102]

4.3 AKA Mechanism

The SN invokes the procedure by selecting the next unused authentication vector from the ordered array of authentication vectors in the SN database. Authentication vectors in a particular node are used on a first-in/first-out basis. The SN sends to the USIM the random challenge RAND and the corresponding authentication token AUTN from the selected authentication vector.

Actions on USIM

Upon receipt of RAND and AUTN the user, resp. the USIM, proceeds as shown in figure 3.

[5] A typical value for *n* could be five.

[6] The AMF serves to define operator-specific options in the authentication process, e.g. the use of multiple authentication algorithms or a limitation of key lifetime.

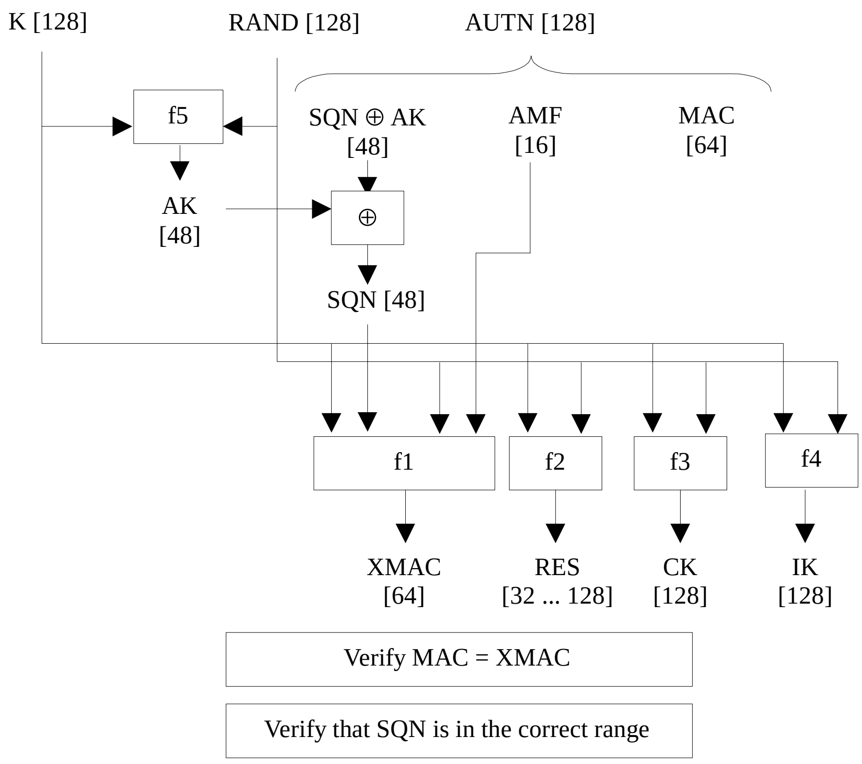

Fig. 3: Authentication function in the USIM [TS 33.102]

The USIM first computes the anonymity key AK = $f5_K$ (RAND) and retrieves the sequence number SQN = (SQN ⊕ AK) ⊕ AK.

Next the USIM computes XMAC = $f1_K$ (SQN || RAND || AMF) and compares this with MAC which is included in AUTN. If they are different, the user sends *user authentication reject* back to the VLR/SGSN with an indication of the cause and the user abandons the procedure.

If the MAC verification was successful, the USIM verifies that the received sequence number SQN is in the correct range (cf. section 4.4).

If the USIM considers the sequence number to be not in the correct range (cf. section 4.4), it sends a *synchronisation failure* message back to the VLR/SGSN including (integrity-protected) information about an acceptable sequence number, and abandons the procedure. The SN then requests fresh authentication vectors from the HE by transferring the *synchronisation failure* message back to the HE.

If the sequence number is considered to be in the correct range however, the USIM computes RES = $f2_K$ (RAND) and includes this parameter in a *user authentication response* back to the SN. Finally the USIM computes the cipher key CK = $f3_K$ (RAND) and the integrity key IK = $f4_K$ (RAND). Upon receipt of the *user authentication response* the SN compares RES with the expected response XRES from the selected authentication vector. If XRES equals RES then the authentication of the user has passed. The SN also selects the appropriate cipher key CK and integrity key IK from the chosen authentication vector.

Protocol Goals Achieved

The protocol achieves the obvious goals of agreeing on symmetric keys CK, IK for ciphering and integrity protection between USIM and Serving Network, and of authenticating the user towards the Serving Network.

In addition, by verifying the MAC included in AUTN, the user is ensured that the random challenge RAND sent by the SN has in fact been generated by the user's HE, and that the SN is trusted by the HE to deal correctly with authentication vectors.

Moreover, although CK and IK are not derived from SQN, by checking that SQN lies in the correct range the user verifies the freshness of the keys CK and IK derived from RAND, because SQN and RAND are jointly integrity protected by the MAC. If SQN is accepted the user trusts his HE that the corresponding RAND was generated randomly and has not been used before. This is considered a minor issue, because there is already a strong trust relationship between user and HE, the latter being in possession of the user's individual key K.

4.4 Sequence Number Management

By Sequence Number Management we mean the rules for generating sequence numbers in the HE and the corresponding rules by which the USIM decides whether to accept a transferred sequence number to be in the correct range or not. Since the USIM is configured and distributed by the HE, sequence number management schemes need not be standardised, but can be operator-specific. However, the standard [TS 33.102] gives example schemes in an informative annex.

When considering the rules the USIM should follow in the decision about acceptance of the sequence number received from the SN, there are basically two issues to be taken into account:

- An array of authentication vectors may arrive at the SN „out of order“, i. e. the initial ordering of the authentication vectors may be disturbed on their way from the HE to the SN. As the sequence numbers are not visible to the SN (they are concealed by the Anonymity Key AK), the SN cannot restore the original ordering. This may result in an AUTN parameter being sent by a legitimate SN to the USIM containing a sequence number which is smaller than a sequence number received by the USIM before. This situation must not lead to a synchronisation failure.
- Accidental or malicious modification of authentication information in the network must not lead to the USIM reaching a state where it permanently rejects authentication requests from the network because the sequence number has been driven up to its maximum possible value *SEQmax*.

These issues are covered by the mechanisms described in the following subsections.

Array Mechanism

Each time an authentication vector is generated, the HE/AuC allocates a certain index value *IND* for that vector according to suitable rules and includes it in the appropriate part of *SQN*. The index value depends on the number of authentication vectors being sent simultaneously to the SN and may range from 0 to *a*-1 where *a* is the size of the array. A typical value for *a* is 32. The USIM maintains an array of *a* previously accepted sequence numbers: SEQ_{MS} (0), SEQ_{MS} (1),…, SEQ_{MS} (*a*-1). The array is initialised with a sequence number value of zero for each array element. To verify that the received sequence number *SQN* is fresh, the USIM compares the received *SQN* with the sequence number in the array element indexed using the index value *IND* contained in the received *SQN*, i.e. with the array entry SEQ_{MS} (i) where i = *IND* is the index value. Now, if

- $SEQ > SEQ_{MS}$ (i), the USIM shall consider the sequence number to be guaranteed fresh and subsequently shall set SEQ_{MS} (i) to *SEQ*.
- $SEQ \leq SEQ_{MS}$ (i), the USIM shall generate a synchronisation failure message indicating the highest previously accepted sequence number SQN_{MS} anywhere in the array.

By using this array mechanism erroneous synchronisation failures are effectively avoided. It may also be used to avoid unjustified rejection of user authentication requests when authentication vectors from different mobility management domains (circuit and packet switched) are used in an interleaving fashion.

Delta Mechanism

In order to avoid the USIM reaching the maximum sequence number *SEQmax* within its assumed lifetime, the USIM should not accept arbitrary jumps in sequence numbers, but only increases by a value of at most Δ, which means that Δ shall be chosen sufficiently large so that the MS will not receive any sequence number with $SEQ - SEQ_{MS} \geq \Delta$ if the HE works correctly. In [TS 33.102] a value of $\Delta = 2^{28}$ is recommended.

In order to prevent that SEQ_{MS} ever reaches the maximum batch number value *SEQmax* during the lifetime of the USIM, the minimum number of steps $SEQmax / \Delta$ required to reach *SEQmax* must be sufficiently large. For $\Delta = 2^{28}$, this means that about 32,000 successful authentications are needed before *SEQmax* is reached.

5 Security Algorithms

For the UMTS security architecture several security algorithms are needed, which have been standardised by 3GPP. The authentication and key agreement algorithms reside in the USIM and the HE/AuC, which both belong to the same network operator. For those operators who do not want to use or do not have the ability to design proprietary algorithms for AKA, 3GPP recommends to use the standard algorithm. The cipher and integrity algorithms on the other hand reside in the mobile and the SN. Therefore it is mandatory that the standard algorithms are used. Algorithm requirements are defined in [TS 33.105].

5.1 Ciphering / Integrity

The cipher/integrity algorithms are used for encryption/integrity protection between the mobile and the RNC. Integrity protection is performed by computing cryptographic checksums for signalling messages.

Requirements

The following requirements have been stated for the functions f8 (ciphering algorithm) and f9 (integrity algorithm):

- The ciphering function f8 shall be a stream cipher.
- The integrity protection function f9 shall be a MAC function.
- Both functions should be implementable in low power, low gate-count hardware, and also perform well in software.
- There should be no export restrictions on terminals.
- Network equipment should be exportable under licence in accordance with Wassenaar agreement.

General Approach to Design

In accordance with these requirements the following approach was chosen by the design authority ETSI SAGE[7]:

- Use a block cipher as a building block for both algorithms.
- Define the modes of operation for this block cipher according to the f8 and f9 requirements.

[7] ETSI SAGE: Security Algorithm Group of Experts, Special Committee for the development of cryptographic algorithms of the European Telecommunications Standards Institute (ETSI)

Since this block cipher has to have minimal hardware and software complexity, either an existing one meeting both requirements had to be chosen or a custom-made had to be developed. Furthermore there must not be any license fees or other restrictions on IPR.

Since the algorithms had to be specified in just six months 3GPP pragmatically decided to take an existing algorithm as a starting point and customise it wherever necessary [TR 33.901]. The starting point chosen was Mitsubishi’s algorithm MISTY1 designed by Mitsuru Matsui [MISTY]. This algorithm is fairly well studied and possesses some provable security aspects.

Furthermore, it was designed for high speed encryption on hardware platforms as well as for software implementations. In addition its parameter sizes fit the requirements. The customised algorithm is now called KASUMI [TS 35.202]. KASUMI is a 64 bit block cipher with a 128 bit key. The intention was to specify an algorithm that could be published according to Kerckhoff’s principle. The specifications are available on the 3GPP web page.

Design and Analysis of Ciphering/Integrity Algorithms

As usual SAGE was divided into a design and evaluation team. Mitsuru Matsui, the designer of MISTY1 joined the design team, the evaluation team was joined by additional evaluators from Nokia, Ericsson and Motorola.

Additionally an external evaluation was done by three teams of renowned cryptologists. The overall report from all evaluation teams is publicly available and confirms the fulfilment of the requirements [TR 33.909].

Design of the Stream Cipher f8

The main task with the design of f8 was to make a stream cipher out of the block cipher KASUMI [TS 35.201]. There are several standard ways for this task, namely cipher feedback (CFB), counter mode or output feedback (OFB) mode. For the function f8 a combination of OFB and counter mode was used for protection of KASUMI against chosen plaintext attacks and protection against collision attacks, cf. figure 4.

To prevent chosen plaintext attacks the initialisation vector (IV) is encrypted with a different key CK’ = CK ⊕ KM, where KM is a key modifier. This initial encryption is also a protection against collision attacks as

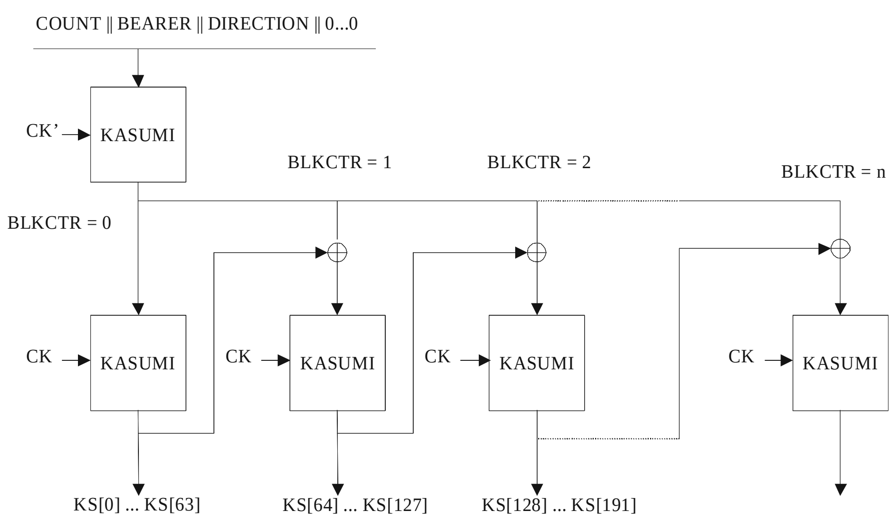

Fig. 4: The UMTS cipher algorithm f8

an attacker cannot freely choose the value which is XORed with the block counter. If it was left out it would be possible to choose two different IVs such that a collision appears at different block counter values (with small probability).

Design of the Integrity Protection Function f9

Algorithm f9 is a CBC-MAC with KASUMI as the underlying block cipher [TS 35.201]. It was constructed in accordance with the standard ISO/IEC 9797-1 (MAC algorithm 2, cf. figure 5).

5.2 Authentication and Key Agreement

The UMTS AKA algorithm consists of seven functions f1, f1*, f2, f3, f4, f5 and f5* (cf. figures 2 and 3). The functions with ‘*’ are used in the resynchronisation event. The standardised algorithm set for these seven functions is called MILENAGE.[8]

Common to all functions is the requirement that it shall be computationally infeasible to derive K from the knowledge of all inputs except K and all outputs.

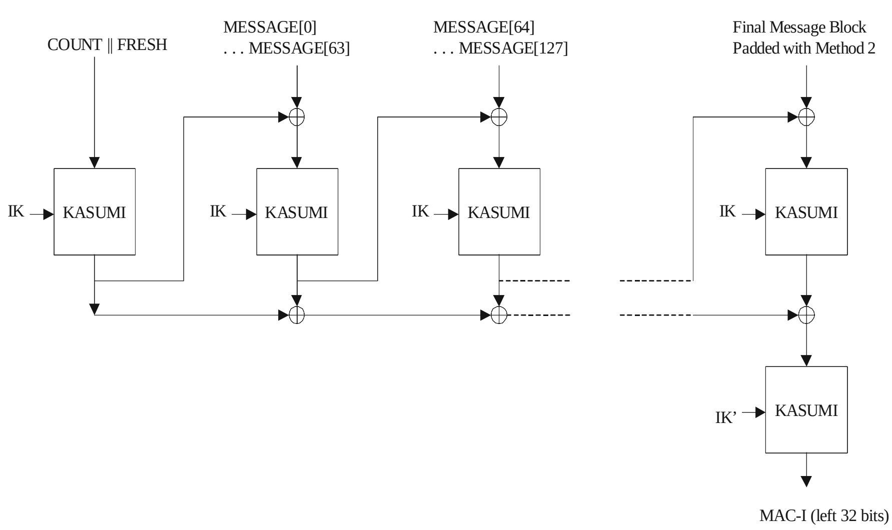

Fig. 5: The UMTS integrity algorithm f9

Design and Analysis of the MILENAGE Algorithm

MILENAGE has been designed by the experienced group that already designed KASUMI, f8 and f9, namely ETSI SAGE, Kaisa Nyberg and Mitsuru Matsui. The evaluation team was joined by Helena Handschuh, Gemplus, as an expert in side channel attacks.

It was required that the AKA algorithm should be built around a kernel function (e.g. a block cipher) which could be replaced by the operator. For MILENAGE, however, a specific kernel had to be chosen, therefore Rijndael was selected.[9]

Even if an operator chooses MILENAGE as his AKA algorithm he can customise it by an operator variant algorithm configuration field OP of 128 bits. The operators are expected to keep their OP values secret. To protect OP against disclosure it should be encrypted offline with the user specific key K and only the cipher text OP_C should be stored on the USIM.

Figure 6 shows the overall design of MILENAGE, where E_K denotes an encryption with Rijndael using K. Beside OP MILENAGE uses constants r1 through r5 and c1 through c5 which might be modified by the operator. They have to be chosen such that the pairs (c*i*, r*i*) are all different and such that c1 has even parity while c2-c5 all have odd parity. The latter requirement ensures that it is impossible to choose RAND, SQN and AMF in a way that the output of the first branch is equal to one of the outputs of the other branches: Assume that OP_C has even parity. Then the input to the second encryption in the first branch has the same parity as $E_K(RAND \oplus OP_C)$, since SQN || AMF || SQN || AMF, OP_C and c1 all have even parity. The inputs to the second encryption in the second branch (and all other branches) have the opposite parity, since OP_C has even whereas c2 (and c3, c4 and c5) have odd parity. Now assume that OP_C has odd parity. Then the input to the second encryption in the first branch has the opposite parity and the inputs to the second encryption in the other branches have the same parity as $E_K(RAND \oplus OP_C)$ by a similar argument.

Functions f1 and f1* are constructed as a standard CBC-MAC with the two input

[8] Algorithm specification and evaluation report are expected to be published soon.

[9] The selection of Rijndael as the AES by the NIST somewhat later confirmed this to be a good choice.

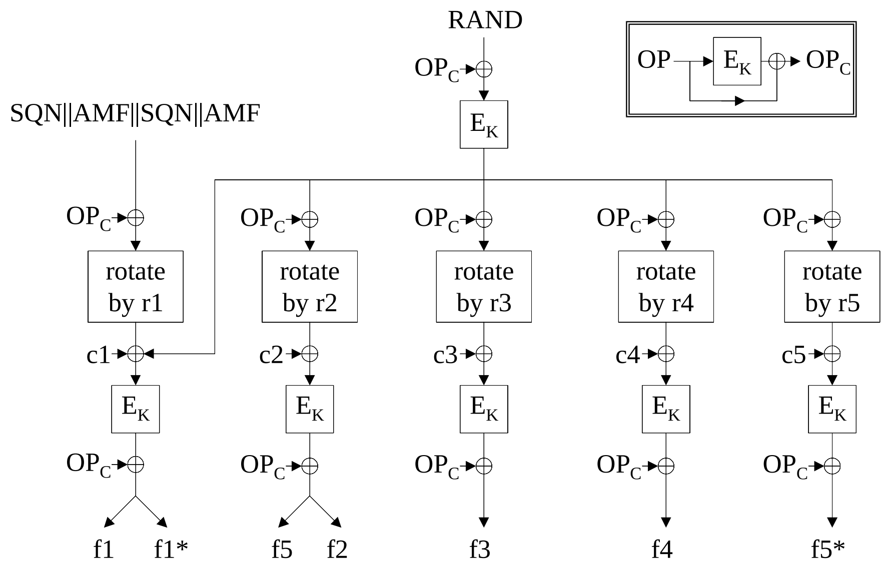

Fig. 6: The MILENAGE example algorithm

blocks $RAND \oplus OP_C$ and $c1 \oplus rot(SQN \parallel AMF \parallel SQN \parallel AMF \oplus OP_C, r1)$. All other functions derive different values from RAND under control of the user specific key K and the operator variant algorithm configuration field OP.

6 Migration from 2G to 3G

Interoperation of 2G and 3G networks will be an important issue in the near future, when both kinds of networks will exist in parallel and need to interwork with each other. Among other things, this implies that 2G subscribers should be able to access 3G networks by means of a 2G SIM card and a 2G/3G dual-mode handset. A guiding principle in the design of 2G/3G interworking mechanisms has therefore been that all practical interworking scenarios shall be supported, i.e. all reasonable combinations of SIM/USIMs[10], 2G/3G handsets, GSM/UMTS Radio Access Network (RAN) and 2G/3G core network nodes.[11]

6.1 Interoperation of 2G and 3G

One obvious problem with 2G/3G interoperation are the different key lengths used in the two systems. After a 3G authentication, the USIM and the SN/HE have agreed on common 128-bit cipher and integrity keys CK and IK. We call this situation the establishment of a *3G security context* between USIM and SN.

After a 2G authentication, on the other hand, only a 64-bit cipher key Kc has been agreed between SIM/USIM and the serving 2G network. In this case, only a *2G security context* with a lower corresponding security level than in the 3G case has been established. In order to cope with this situation, certain conversion functions are needed, that convert (i.e. shorten or blow up) the 3G keys to 2G length and vice versa.

There are two basic scenarios that need to be distinguished and which will be addressed in the following subchapters: We speak of *UMTS Subscriber Roaming* (or USIM Roaming, for short) when a USIM requests access to a 3G or 2G radio access network. Accordingly, *GSM Subscriber Roaming* (GSIM Roaming) is the case when a SIM tries to authenticate in a 3G or 2G access network. There are several subcases to be considered in these basic scenarios, which have been presented elsewhere [3G2000, TR 31.900]. In this contribution, we will focus on the basic mechanisms supporting USIM Roaming and the conversion functions.

6.2 USIM Roaming

The basic assumption to be made here is that 2G entities are not capable of dealing with 3G authentication vectors (with the exception of 2G RAN, which are transparent for 3G authentication parameters RAND, AUTN, RES, but nevertheless can only handle 2G ciphering keys Kc), nor are they able to do any conversion. This means all necessary conversions from 2G to 3G security context and vice versa have to be performed by the involved 3G entities. Figure 7 visualises the possible scenarios and necessary conversions for a USIM inserted into a 3G mobile station requesting access to a 2G or 3G RAN (It is also assumed that the user uses a dual-mode handset, capable of accessing both 2G and 3G RANs).

A 3G security context can be established (i.e. the UMTS AKA described in chapter 4 can be performed), if the HE of the user is 3G **and** the VLR of the SN controlling the RAN, where access is being requested from the USIM, is also 3G . Note that there may be cases where a 3G VLR controls both 3G and 2G RANs (cases A, B in figure 5). However, if a USIM belonging to a 3G HE authenticates at a 3G VLR environment, a 3G security context has been established, regardless of the RAN type.

In the remaining cases, only a 2G security context can be established, either if the user's HE is 2G (cases D – E) **or** if the HE is 3G, but the controlling VLR is 2G only (case C).

In the case of a packet-switched call, there may also occur a change of the anchor VLR during the call (Handover). If this change happens to be from 3G VLR to 2G VLR, the 3G cipher and integrity keys have to be converted to the 2G key cipher Kc in before. In case of a previously established 3G security context, this context is lost and replaced by a 2G security context.

On the user side, the USIM always executes functions f2, f3, f4 to get RES, CK, IK from RAND (cf. 4.3.1). Subsequently, it performs the necessary conversion to 2G parameters SRES, Kc, if access to a 2G RAN is requested (cases B – E).

[10] For UMTS, one SIM application may reside jointly to one or several USIM applications on a single UICC (UMTS Integrated Circuit Card)

[11] In what follows, it is assumed that the reader is familiar with the basic network architecture of GSM and UMTS.

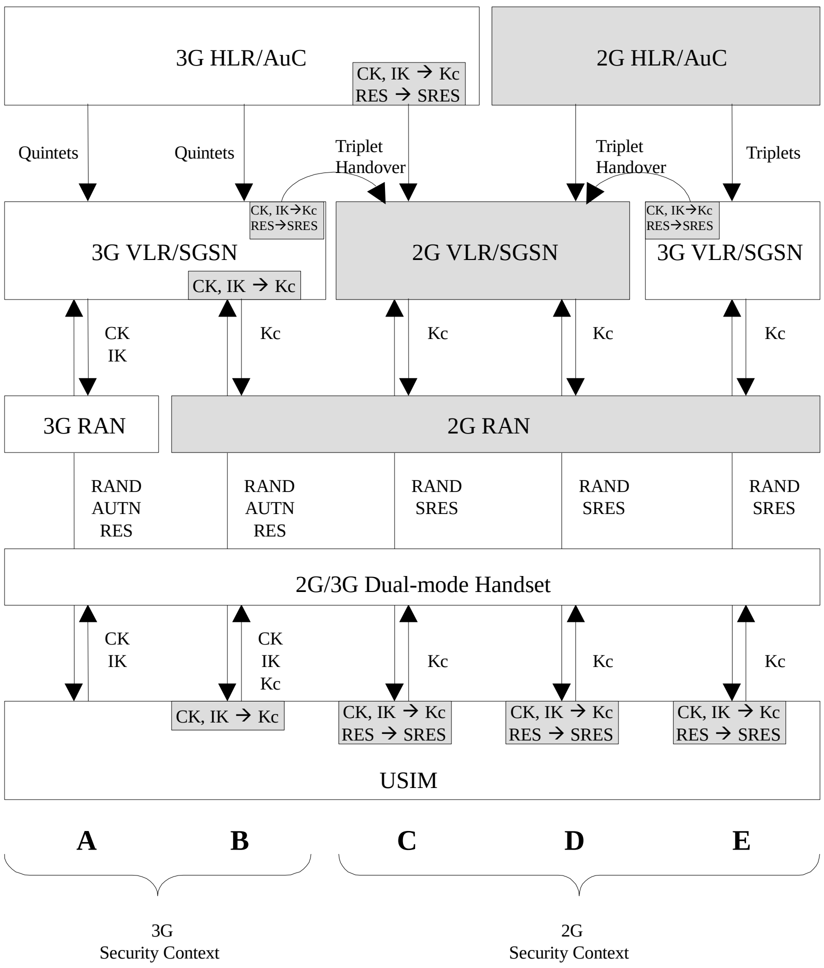

Fig. 7: USIM roaming

6.3 GSIM Roaming

GSIMs (i.e. a SIM application running on a 3G UICC or a traditional GSM SIM) may be inserted into both 2G or 3G mobile stations. In each possible scenario (cf. [3G2000, TR 31.900]) only the 2G AKA is performed. If necessary, conversion of the 2G to 3G authentication data are performed by the involved 3G entities. The user gets 2G security level in every case, also when a 3G mobile station is used and access to a RAN is requested, even if in this case the security level is slightly higher than for „pure" GSM, since the 3G RAN invokes integrity protection by the derived integrity key IK.

6.4 Conversion Functions

Standardised conversion functions c_N are necessary to convert 2G to 3G security parameters and vice versa. c1 – c3 convert 3G security parameters RAND, RES_{3G}, CK, IK into 2G authentication triplets RAND, RES_{2G}, Kc. Conversion function c4 – c5 convert 2G cipher keys Kc into 3G cipher and integrity keys CK, IK. Since these functions need to be implemented in various network nodes and mobile equipment manufactured by different vendors, they need to be standardised. It should be noticed that these functions are publicly available and do not bear any cryptographic significance [TS 33.102].

- **Conversion function c1** converts the 3G random challenge into the 2G random challenge. However, 3G and 2G random challenges will be equal, so that c1 is just the identity function:
  $$RAND_{2G} = c1(RAND_{3G}) = RAND_{3G}$$
- **Conversion function c2** converts a 3G expected authentication response XRES into a 2G expected authentication response RES (done in the HE/AuC or the VLR) or a 3G authentication response RES into a 2G authentication response SRES (done in the USIM):
  $$RES_{2G} = c2(XRES_{3G}),$$
  $$SRES_{2G} = c2(RES_{3G})$$
- **Conversion function c3** converts the 128 bit 3G ciphering and integrity protection keys CK and IK into the 64 bit 2G ciphering key Kc. This function is applied in the HE/AuC or VLR and in the USIM.
  $$Kc = c3(CK, IK)$$
- **Conversion function c4** converts a 64 bit 2G Kc into a 128 bit 3G CK. This function is applied in the ME and in the VLR.
  $$CK = c4(Kc)$$
- **Conversion function c5** converts a 64 bit 2G Kc into a 128 bit 3G IK. This function is applied in the ME and in the VLR.
  $$K = c5(Kc)$$

6.5 Compatibility Conditions for Authentication

In the future combination cards (UICC with SIM and USIM application) are expected. This seems at least for incumbents the best choice to migrate to 3G. In general there are two possibilities to link SIM/USIM applications and 2G/3G subscriptions (same/separate IMSI per SIM/USIM application). In case of assigning the same IMSI (International Mobile Subscriber Identity) for both 2G and 3G networks, this requires to use the same authentication key (K_{GSM} = Ki = K = K_{UMTS}). The need for standardised conversion functions implies certain compatibility requirements on the interplay of authentication and key agreement functions A3, A8 the operator uses for GSM on a 2G/3G combination card and f2, f3, f4 the operator uses for UMTS on a combination card. Namely, in order to ensure that GSM authentication data (SRES, Kc) derived in the USIM from a UMTS authentication vector are identical to the data computed in a 2G HLR or converted in a 3G HLR/VLR, the authentication and key agreement functions A3/f2 and A8/(f3,f4) must satisfy the conditions

$$SRES_{2G} = A3_{Ki}(RAND) = c2(f2_K(RAND)) = c2(RES_{3G})$$

and

$$Kc = A8_{Ki}(RAND) = c3(f3_K(RAND), f4_K(RAND)) = c3(CK,IK).$$

7 Network Domain Security

The core network of mobile radio systems is the part of the network which is independent of the radio interface technology. It is used for transporting user data as well as signalling commands needed to ensure smooth operation of the overall system. Although eavesdropping or modifying signalling messages in the core network would have serious consequences for mobile radio systems because sensitive authentication data of mobile subscribers are transported across the core network. In 2G mobile systems like GSM the core network is not secured. If an intruder succeeded in eavesdropping on these authentication data, serious impersonation attacks or eavesdropping on user traffic on the air interface might result. To date, however, no such attacks on GSM core networks are known.

One reason for not securing the core network of 2G mobile systems was the initial design goal of making 2G mobile systems only as secure as fixed network connections which meant in effect that security features of 2G systems were basically constrained to the most vulnerable part of the network, the radio interface. A second, equally important, reason was that the core network was considered „closed“, i.e. without external interfaces which might be used by attackers. GSM uses the SS7 (Signalling System No. 7) protocol stack with MAP (Mobile Application Part) as an application on top for its signalling messages. Due to its complexity and the limited availability of implementations, detailed knowledge of SS7 is confined to telecoms insiders which helps to reduce the potential risk.

All these assumptions, however, do no longer hold for 3G systems such as UMTS: with the introduction of IP-based transport to most, if not all, interfaces of the UMTS network reference model new vulnerabilities of the core network as well as new potential threats directed towards the core network from the outside have to be taken into account. In addition to building, and managing, their own „private“ transport networks, operators also have the technical possibility to rent the transport capacity required between any two nodes of the reference model from virtually any ISP offering real-time transport services.

For these reasons, the security features of UMTS will also cover the core network. Protection is provided for only signalling messages, as attacks on these seem to pose the greatest risk.

In the present contribution, only a very brief overview of the basic concepts on 3G core network signalling security is given. More detailed information can be found in [CMS2001].

7.1 Securing Protocols of the Core Network

In GSM, the Mobile Application Part (MAP) and the CAMEL Application Part (CAP) are used for signalling between the various network elements. MAP/CAP run as an application on top of the SS7 protocol stack. 3GPP has based its specifications for UMTS, e.g. [TS 29.002], on an evolved GSM core network, thereby retaining the basic core network architecture of GSM and its protocols. It is envisaged that the SS7-based transport stack for MAP will be gradually replaced by an IP-based transport stack in the UMTS core network. But there will be quite a long period where nodes supporting MAP/CAP over IP will have to communicate with nodes supporting MAP or CAP over SS7, e.g. nodes in a 2G network. Because of the difficulties involved with this situation, it was decided to secure these „legacy“ protocols at the application layer, irrespective of their transport stack.

In addition to MAP and CAP, there will be purely IP-based protocols like GTP (GPRS Tunneling Protocol). It is quite natural to secure these „native“ IP-based protocols at the IP-layer by deploying IPSec.

7.2 Securing Legacy Protocols

Secured MAP messages consist of a MAP message header, a security header and the protected payload that is the result of applying the corresponding protection mode (i.e. desired level of protection: no protection, integrity, confidentiality and integrity) to the original MAP message payload. The exact message format of secured MAP messages can be found in [CMS2001]. In all three protection modes, the security header is transmitted in cleartext. Among other pieces of information, it contains the sending network identity and the Security Parameter Index (SPI), an arbitrary 32-bit value that is used in combination with the sending network identity to uniquely identify a MAP Security Association (SA). Just as IPSec SAs do, the MAP SAs specify the parameters needed to protect the communication between two network elements over the so-called Z_C interfaces. In contrast to IPSec SAs, MAP SAs are network-wide SAs, i.e. the SAs are valid for a specific pair of networks for a certain period. Network-wide SAs are necessary because, in general, it is not possible for a sending MAP network entity in a network A to determine the address of the receiving network entity in a network B. This is due to the particularities of routing for MAP over SS7 which is done by intermediate gateways based on the user identity IMSI.

The proposed mechanism for key management consists of a two-tiered architecture with a new network element, the Key Administration Center (KAC), in every network. KACs communicate with each other over the IP-based Z_A interface and negotiate MAP SAs by using IKE, the Internet Key Exchange mechanism. The KACs then distribute the MAP SAs further to the network elements.

7.3 Securing IP-based Communication

All IP communication that requires to be protected is routed in a hop-by-hop fashion through protected tunnels. To support this approach between different networks, a new logical entity called SEcurity Gateway (SEG) has been added to the architecture. SEGs operate at the border of a network, providing IP security for inter-network IP traffic. SEGs establish and maintain IPSec tunnels with any network entity of their own network that uses this SEG to secure intra-network IP traffic destined for other networks. Similarly, IPSec tunnels may be established between two network elements residing in the same network. Depending on the network configuration the SEG entities support a single, uniform IPSec tunnel to another network that tunnels all types of IP communication, or they establish several tunnels applying different security services for different protocols, ports or even hosts. Policy information for these secure tunnels has to be exchanged in advance, as part of the roaming agreement, between the operator’s SEG entities.

With respect to key management, the initial idea within the 3GPP security group was to design a unified key management where the KACs also provide security associations for IPSec to support the IP-based network elements. This approach proved to be difficult, because, for IPSec, individual SA need to be established peer-to-peer between two network elements. On the other hand, SA negotiation in the IPSec standards is intended not to take place remotely by a third party, but directly between the IPSec peers that use these SAs subsequently. As a consequence, 3GPP decided not to use the KACs for IPSec SA negotiation, but to support direct SA negotiation between IPSec hosts.

Between different networks, it is mandated that all IP traffic passes an SEG at each network border. SAs between two networks are therefore only required between specific SEGs that connect the networks. These SEGs support the Internet Key Exchange protocol IKE, which is the default key management protocol for IPSec. With IKE, the SEGs can negotiate their IPSec SAs directly. Standard IPSec implementations can be used here. For further details concerning the IPSec configuration, see [CMS2001].

8 Conclusion

The security architecture of the 3G mobile radio system UMTS has now reached a reasonably stable state. Since some open issues remain in the area of Internet Multimedia Security (IMS), this field has not been discussed in the present contribution. However, with the security features now being on the table, it can be concluded that UMTS security will improve greatly on 2G security and provide its users with a level of security formerly unknown in public telephony systems. In order to reach this goal, completely new security features were included into the UMTS security architecture, such as the authentication of HE towards the USIM. But also the design and evaluation process of the UMTS security algorithms was based on a new philosophy: After having its algorithms evaluated by internal and external experts, 3GPP made all specifications and evaluation reports publicly available and open to public scrutiny.[12] This will not only further improve on the overall security of the UMTS system, but will also foster public trust in its security, thereby contributing to the commercial success of the 3G mobile radio system UMTS.

[12] 3GPP specifications are available from www.3gpp.org.

References

[3G2000] Pütz, S.; Schmitz, R.: *Secure Interoperation between 2G and 3G Mobile Radio Networks.* Proc. of 3G2000. First International Conference on Mobile Communications Technologies. IEE CP 471, 2000.

[CMS2001] Horn, G.; Kröselberg, D.; Pütz, S.; Schmitz, R.: *Security for the core network of third generation mobile system.* Proc. of Communications and Multimedia Security CMS 2001. To appear in 2001.

[DuD1996] Pütz, S.: *Zur Sicherheit digitaler Mobilfunksysteme.* Datenschutz und Datensicherheit (DuD), 6/97, S. 321-327.

[DuD2001] Pütz, S.; Schmitz, R.: UMTS Sicherheitsdienste. Datenschutz und Datensicherheit (DuD), 4/2001, S. 205-207.

[GSM 02.09] ETSI Standard GSM 02.09: GSM Security Aspects.

[GSM 03.20] ETSI Standard GSM 03.20: GSM Security Architecture.

[ISSE2000] Horn, G.; Howard, P.: *Review of Third Generation Mobile System Security.* Proc. Information Security Solutions Europe (ISSE) 2000.

[MISTY] Matsui, M.: *New Block Encryption Algorithm MISTY.* Proc. of 4th International Workshop on Fast Software Encryption. Lecture Notes in Computer Science 1267, pp. 54-68, Springer, 1997.

[TR 31.900] 3GPP Technical Report 31.900: *SIM/USIM Internal and External Interworking Aspects.*

[TR 33.901] 3GPP Technical Report 31.901: *Criteria for cryptographic algorithm design process.*

[TR 33.902] 3GPP Technical Report 33.902: *Formal analysis of the 3G authentication protocol.*

[TR 33.909] 3GPP Technical Report: *Report on the evaluation of 3G standard confidentiality and integrity algorithms.*

[TS 29.002] 3GPP Technical Specification 29.002: *Mobile Application Part (MAP) for UMTS.*

[TS 33.102] 3GPP Technical Specification 33.102: *UMTS Security architecture.*

[TS 33.105] 3GPP Technical specification 33.105: *UMTS Cryptographic algorithm requirements.*

[TS 33.120] 3GPP Technical specification 33.120: *UMTS Security principles and objectives.*

[TS 35.201] 3GPP Technical Specification 35.201: *Design of the UMTS Cipher and Integrity Functions.*

[TS 35.202] 3GPP Technical Specification 35.202: *Design of the KASUMI Block Cipher.*

[VIS2001] Martin, T.; Pütz, S.; Schmitz, R.: *On the Security of the UMTS System.* Accepted for Verlässliche IT-Systeme (VIS) 2001.