UMTS security

by K. Boman, G. Horn, P. Howard and V. Niemi

The Universal Mobile Telecommunications System (UMTS) is one of the new 'third generation' mobile cellular communication systems. UMTS builds on the success of the 'second generation' GSM system. One of the factors in the success of GSM has been its security features. New services introduced in UMTS require new security features to protect them. In addition, certain real and perceived shortcomings of GSM security need to be addressed in UMTS. This paper surveys the major security features that are included in the first releases of the UMTS standards.

1 Introduction

The Universal Mobile Telecommunications System (UMTS) is one of the new 'third generation' (3G) mobile cellular communication systems being developed within the framework defined by the ITU and known as IMT-2000[1]. UMTS builds on the capability of today's mobile technologies by providing increased capacity, data capability and a greater range of services using a new radio interface standard called UMTS Terrestrial Radio Access (UTRA)[2].

The basic radio, network and service parameters of the UMTS system were defined by the European Telecommunications Standards Institute (ETSI) in early 1998[3]. ETSI developed the extremely successful second generation GSM (Global System for Mobile communications) standard, which is used by over 650 million customers world-wide and accounts for approximately 70% of the wireless communications market[4]. An important characteristic of UMTS is that the new radio access network will be connected to an evolution of the GSM core network.

To help build on the global success of GSM, the UMTS standards work in ETSI was transferred in 1998 to a global partnership of regional standards bodies called 3GPP (3rd Generation Partnership Project)[5]. A separate partnership of standards bodies, known as 3GPP2[6], is developing another third generation mobile cellular system based on a different 3G radio interface standard called CDMA2000 and a core network that is evolved from the North American ANSI-41 standard.

One of the aspects of GSM that has played a significant part in its global appeal is its set of security features. UMTS security builds on the success of GSM by providing new and enhanced security features. This paper surveys the main security features that are included in the first releases of the UMTS standards.

2 Building on GSM security

Security has always been an issue for mobile phones. Many of the 'first generation' analogue mobile phone systems were susceptible to abuse. For example, it was possible to eavesdrop on the analogue radio path and thereby listen to other people's calls, or to reprogram the identities of mobile phones such that the cost of calls made using them appeared on another customer's bill. It was against this background of user traffic eavesdropping and cloning fraud that the 'second generation' GSM system was designed.

GSM was designed to prevent cloning and to be no more vulnerable to eavesdropping than fixed phones. It addresses these goals by providing user-related security features for *authentication*, *confidentiality* and *anonymity*[7]. The authentication feature is intended to allow a GSM network operator to verify the identity of a user such that it is practically impossible for someone to make fraudulent calls by masquerading as a genuine user. Confidentiality protects the user's traffic, both voice and data, and sensitive signalling data, such as dialled telephone numbers, against eavesdropping on the radio path. The anonymity feature was designed to protect the user against someone who knows the user's international mobile subscriber identity (IMSI) from using this information to track the location of the user, or to identify calls made to or from the user by eavesdropping on the radio path.

The most novel feature of GSM security is the use of a smart card, known as the subscriber identity module (SIM). The SIM contains all the identification and security-related data that the subscriber needs to make or receive a call. It is, in effect, a portable security module, personalised for the subscriber. The SIM can be used to access services in any network with which the subscriber's home network has a roaming agreement. During roaming the subscriber's home network provides all the data needed by the serving network to operate the security features without revealing any of the sensitive security data stored in the subscriber's SIM.

The GSM security features have addressed to a very large extent the needs of operators and the aspirations of users. UMTS security builds on the success of GSM by retaining the security features that have proved to be needed and that are robust. As in GSM, a smart card is

Abbreviations

3G	= 3rd Generation
3GPP	= 3rd Generation Partnership Project
AES	= Advanced Encryption Standard
AH	= Authentication Header
AKA	= Authentication and Key Agreement
API	= Application Programming Interface
AuC	= Authentication Centre
CBC-MAC	= Cipher Block Chaining Message Authentication Code
ESP	= Encapsulating Security Payload/ Encapsulated Security Protocol
ETSI	= European Telecommunications Standards Institute
GSM	= Global System for Mobile communications
HE	= Home Environment
HSS	= Home Subscriber System
HTTP	= HyperText Transfer Protocol
I-CSCF	= Interrogating Call Session Control Function
IETF	= Internet Engineering Task Force
IK	= Integrity Key
IKE	= Internet Key Exchange
IMEI	= International Mobile Equipment Identity
IMPI	= IP Multimedia Private Identity
IMPU	= IP Multimedia Public Identity
IMS	= IP Multimedia Subsystem
IMSI	= International Mobile Subscriber Identity
ISIM	= IMS Subscriber Identity Module
IP	= Internet Protocol
ITU	= International Telecommunication Union
KAC	= Key Administration Centre
MAC	= Medium Access Control
MAC(-I)	= Message Authentication Code (for data Integrity)
MAP	= Mobile Application Part
MAPSEC	= MAP Security Protocol
ME	= Mobile Equipment
MS	= Mobile Station
PAMPAS	= Pioneering Advanced Mobile Privacy And Security
P-CSCF	= Proxy Call Session Control Function
PIN	= Personal Identity Number
PKI	= Public Key Infrastructure
PS	= Packet Switched
RLC	= Radio Link Control
RNC	= Radio Network Controller
RRC	= Radio Resource Control
RTP	= Realtime Transport Protocol
SA	= Security Association
S-CSCF	= Serving Call Session Control Function
SDP	= Session Description Protocol
SEG	= Security Gateway
SGSN	= Serving GPRS Support Node
SHAMAN	= Security for Heterogeneous Access in Mobile Applications and Networks
SIM	= subscriber identity module
SIP	= Session Initiation Protocol
S/MIME	= Secure/Multipurpose Internet Mail Extensions
SN	= Serving Network
SPI	= Security Parameters Index
SS7	= Signalling System Number 7
TCP	= Transmission Control Protocol
TLS	= Transport Layer Security
TS	= Technical Specification
UA	= User Agent
UDP	= User Datagram Protocol
UMTS	= Universal Mobile Telecommunications System
USIM	= Universal Subscriber Identity Module
UTRA(N)	= UMTS Terrestrial Radio Access (Network)
VLR	= Visitor Location Register

used in UMTS to store all the identification and security-related data that the subscriber needs to make or receive a call.

Although GSM security has been very successful, an objective of the UMTS security design was to improve on the security of second generation systems like GSM by correcting real and perceived weaknesses. Some of the issues that have had an impact on the design of the UMTS access security architecture are listed below[8,9].

- The currently used GSM cipher algorithms (used to provide confidentiality) are not published along with the bulk of the GSM standards. Instead, the GSM Association controls the distribution of the algorithm specifications. The decision not to make the algorithms available for peer review has received some criticism, with hindsight, from the academic world. However, it must be recognised that GSM security was designed at a time when the controls on the export and use of cryptography were much tighter. The regulatory situation was considerably relaxed in the late 1990s, which led 3GPP to adopt a more open approach to the design of the UMTS algorithms and to publish the algorithm specifications together with the rest of the UMTS standards.
- Unlike the cipher algorithm, the GSM and UMTS authentication algorithms do not need to be standardised and operators are free to design or select their own. In GSM an example algorithm was not included in the standards. This resulted in some operators using an algorithm, known as COMP-128, that has been recognised to be vulnerable to cryptographic attack. After this attack was published on the Internet, the GSM Association made a replacement algorithm available. To help avoid inadequate algorithms being used in UMTS, an example algorithm called MILENAGE[10] has been included in the standards for use by operators who do not wish to design their own.
- The strength of the cipher algorithm depends in part on the length of the cipher key. In GSM the cipher key is transported as a 64 bit structure. However, in practice the top 10 bits of the cipher key are set to zero to reduce

the effective key length to 54 bits. This was due to the regulatory controls that were in force when GSM was designed. As these controls have been relaxed it is now possible for GSM to use full-length 64 bit keys. Creating a longer key than this for GSM is much more complex because it would require the ciphering algorithm to be replaced and the signalling protocols to be upgraded to support the longer key. UMTS required a new ciphering mechanism anyway so the opportunity was taken to increase the cipher key length to 128 bits, which should provide a good level of security for many years to come.

- GSM was not explicitly designed to protect against active attacks on the radio path, because they would require an attacker to masquerade as a GSM network (so-called 'false base station attacks'). These attacks were considered to be too expensive to mount compared to other attacks. However, as mobile cellular services become more widespread, the cost and availability of equipment that may be used to masquerade as a base station make such attacks more likely. Although GSM already provides some protection against certain types of false base station attack, a more thorough threat analysis was performed during the UMTS design phase. This has led to the development of new security features, which are explicitly designed to counteract false base station attacks.
- For GSM circuit-switched services, user traffic and sensitive signalling information are protected on the GSM radio path between the mobile and the base station using a ciphering algorithm. While this protects communications on the most vulnerable radio path, an opportunity was taken in UMTS to extend ciphering further back into the network. This allows more links within the radio access network to be automatically protected, including potentially vulnerable microwave links that may be used to connect base stations to the fixed part of the network.

3 UMTS access security

The access security features in UMTS are a superset of those provided in GSM. They are specified in TS 33.102[11], which is included in the first major release of the 3GPP specifications, known as Release 99. The security features in UMTS that are new compared to GSM are introduced to correct the real and perceived weaknesses of GSM security explained in Section 2 of this paper. The UMTS access security standards, in particular the new authentication mechanism, are based on research work conducted by the European Union funded USECA project[12].

Entity authentication

UMTS provides mutual authentication between the UMTS subscriber, represented by a smart card application known as the USIM (Universal Subscriber Identity Module), and the network in the following sense:

- Subscriber authentication: the serving network corroborates the identity of the subscriber.
- Network authentication: the subscriber corroborates that he is connected to a serving network that is authorised, by the subscriber's home network, to provide him with services; this includes the guarantee that this authorisation is recent.

It should be noted here that the concept of authentication, in general, has many subtle aspects[13,14].

Signalling data integrity and origin authentication

The following security features are provided with respect to integrity of data on the network access link:

- Integrity algorithm agreement: the mobile station (MS) and the serving network (SN) can securely negotiate the integrity algorithm that they use.
- Integrity key agreement: the MS and the SN agree on an integrity key that they may use subsequently; this is realised as part of the protocol which also provides entity authentication.
- Data integrity and origin authentication of signalling data: the receiving entity (MS or SN) is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending entity (SN or MS) and that the data origin of the signalling data received is indeed the one claimed. The use of the integrity feature for signalling data is mandatory.

This security feature has no equivalent in GSM. It provides protection against false base station attacks as the origin of signalling messages required to set up a communication with a mobile can now be authenticated by the mobile.

User traffic confidentiality

The following security features are provided with respect to confidentiality of data on the network access link:

- Ciphering algorithm agreement: the MS and the SN can securely negotiate the ciphering algorithm that they use.
- Cipher key agreement: the MS and the SN agree on a cipher key that they may use subsequently; this is realised as part of the protocol that also provides entity authentication.
- Confidentiality of user and signalling data: neither user data nor sensitive signalling data can be overheard on the radio access interface.

This security feature is the same as in GSM, but the entities between which protection is afforded are different. In UMTS, the protection extends to the radio network controller (RNC), so that microwave links between the base stations and the RNC are also covered.

User identity confidentiality

The following security features related to user identity confidentiality are provided:

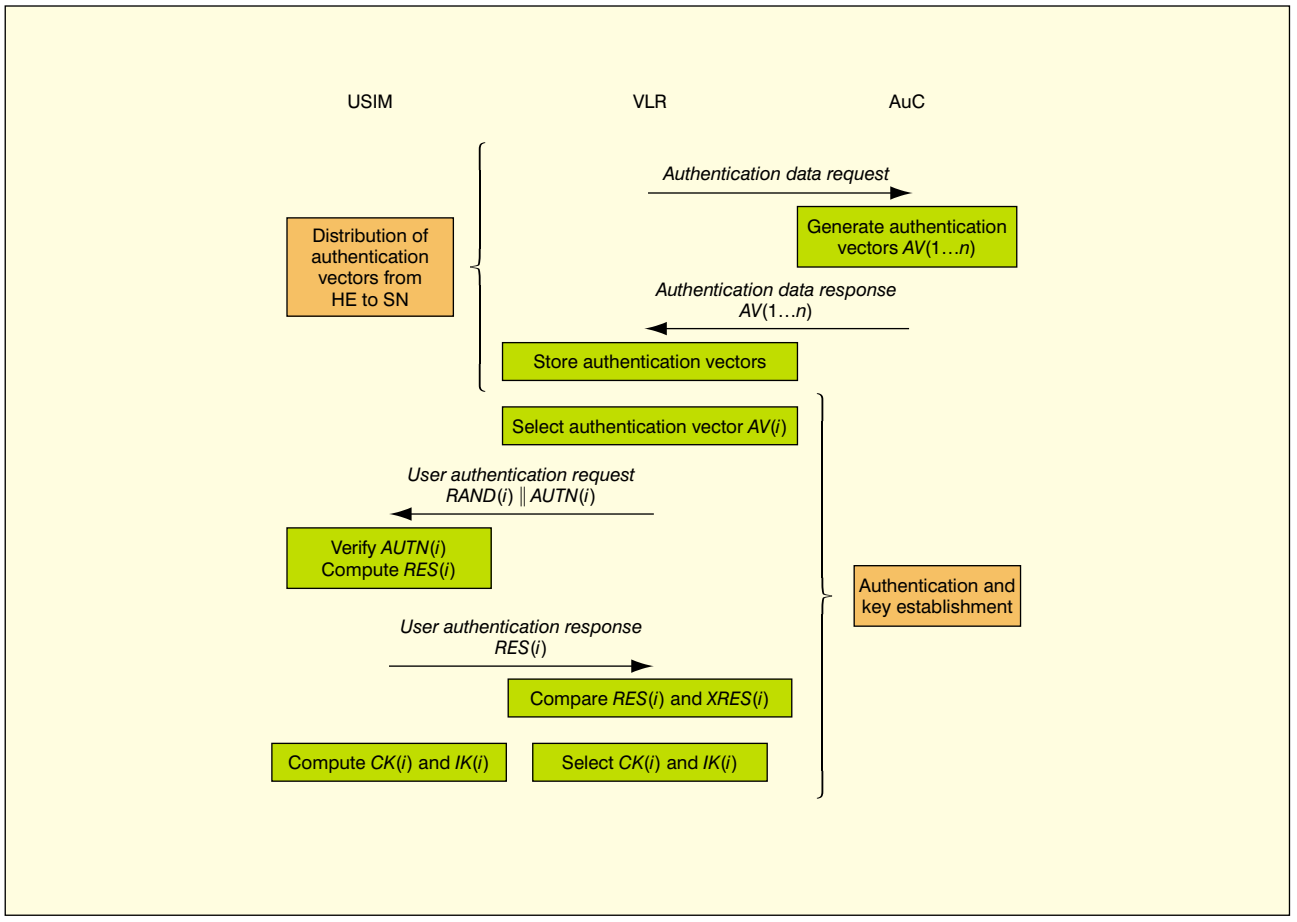

Fig. 1 Overview of authentication and key agreement

- User identity confidentiality: the permanent user identity (IMSI) of a user to whom services are delivered cannot be eavesdropped on the radio access link.
- User location confidentiality: the presence or the arrival of a user in a certain area cannot be determined by eavesdropping on the radio access link.
- User untraceability: an intruder cannot deduce whether different services are delivered to the same user by eavesdropping on the radio access link.

To achieve these objectives, the user is normally identified on the radio access link by a temporary identity by which he is known at the serving network. To avoid user traceability, which may lead to the compromise of user identity confidentiality, the user should not be identified for a long period by means of the same temporary identity. In addition it is required that any signalling or user data that might reveal the user's identity is ciphered on the radio access link. These features are identical to those provided in GSM. They protect against passive attacks, but not against active attacks.

Mobile equipment identification

This feature is the same as in GSM. In certain cases, the serving network may request the mobile station to send its international mobile equipment identity (IMEI). Neither GSM nor UMTS provide a method for authenticating the mobile equipment identity. This is largely due to the complexity of designing and implementing a robust system. This means that any network features that are based on the IMEI, e.g. the barring of stolen phones, relies on the terminal providing the genuine IMEI to the network. The standards therefore impose requirements on terminals to protect the integrity of the IMEI so that it cannot be tampered with or reprogrammed.

User-to-USIM authentication

This feature is the same as in GSM. It provides the property that access to the USIM is restricted until the USIM has authenticated the user. Thereby, it is ensured that access to the USIM can be restricted to an authorised user or to a number of authorised users. To accomplish this feature, user and USIM must share a secret (e.g. a personal identity number, or PIN) that is stored securely in the USIM. The user gets access to the USIM only if he/she proves knowledge of the secret.

USIM–terminal link

This feature ensures that access to a terminal or other user equipment can be restricted to an authorised USIM. To this end, the USIM and the terminal must share a secret that is stored securely in the USIM and the terminal. If a USIM fails to prove its knowledge of the secret, it will be denied access to the terminal. This feature is the same as for GSM. It is also known as 'SIM-lock'.

Secure messaging between the USIM and the network

The USIM Application Toolkit provides the capability for operators or third-party providers to create applications that are resident on the USIM (similar to SIM Application Toolkit in GSM). There exists a need to secure messages, which are transferred over the network to applications on the USIM, with the level of security chosen by the network operator or the application provider.

Visibility of security

Although in general the security features should be transparent to the user, for certain events and according to the user's concern, greater user visibility of the operation of security features should be provided. Therefore, in UMTS, a ciphering indicator is mandatory. It shows the user whether the confidentiality of user data is protected on the radio access link, in particular when non-ciphered calls are set-up. However, the ciphering indicator can be deactivated by the operator who issues the USIM by setting the appropriate bit in the USIM.

Mutual authentication and key agreement between user and network

The design of the authentication and key agreement (AKA) protocol for UMTS reflects the results of an analysis of the threats and risks in GSM. It was guided by the principle that the compatibility with GSM should be maximised and the migration from GSM to UMTS, and the handover between GSM and UMTS access networks, should be made as easy as possible. In particular, the changes to the GSM core network should be minimised.

The main changes with respect to the GSM authentication and key agreement protocol are:

- The challenge is protected against replay by a sequence number and it is also 'signed' (integrity-protected). This means that old authentication data intercepted by an attacker cannot be reused.
- The AKA generates an integrity key in addition to a ciphering key. This integrity key is used to protect the integrity of the signalling data between the MS and the RNC.

In the following an overview of how the UMTS AKA protocol works is given.

Prerequisites

There are three parties communicating in the protocol: the authentication centre (AuC) in the home environment (HE) of the user, the visitor location register (VLR) in the serving network (SN) and the user, represented by his universal subscriber identity module (USIM). In the case of the packet-switched (PS) domain of UMTS, the role of the VLR is taken by the serving GPRS support node (SGSN).

The UMTS AKA protocol is of the secret-key type. One secret key, the authentication and key agreement key K, is shared by two parties, the AuC and the USIM.

The following trust relations are assumed: the user trusts the HE in all respects concerning this protocol. The HE trusts the SN to handle authentication information, sent by the HE to the SN, securely. The HE distributes authentication information only to SN entities it trusts to provide services to the HE's users. The SN trusts the HE to send correct authentication information and pay for the services provided by the SN to the HE's users. The SN accepts authentication information only from entities it trusts.

It must be further assumed for the protocol to be secure that the intra-system interfaces linking the SN to the HE, and linking SNs, are adequately secure. This security is provided by network domain security, as described in Section 4.

Procedures

Authentication and key agreement (Fig. 1) consists of two procedures. First, the HE distributes authentication information to the SN. Second, an authentication exchange is run between the user and the SN. The authentication information consists of the parameters necessary to carry out the authentication exchange and provide the agreed keys.

Fig. 1 shows that, after receiving an authentication data request, the HE generates an ordered array of n authentication vectors. Each authentication vector consists of five components (and hence may be called a UMTS 'quintet' in analogy to GSM 'triplets'): 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 from the HE to the SN. It is good for n authentication exchanges between the SN and the USIM. In an authentication exchange the SN first selects the next (the i-th) authentication vector from the array and sends the parameters $RAND(i)$ and $AUTN(i)$ to the user. The USIM checks whether $AUTN(i)$ can be accepted and, if so, produces a response $RES(i)$, which is sent back to the SN. $AUTN(i)$ can only be accepted if the sequence number contained in this token is fresh. The USIM also computes $CK(i)$ and $IK(i)$. The SN compares the received $RES(i)$ with $XRES(i)$. If they match, the SN considers the authentication exchange to be successfully completed. The established keys $CK(i)$ and $IK(i)$ will then be transferred by the USIM to the mobile equipment and by the VLR (or SGSN) to the RNC; the keys are then used by the ciphering and integrity functions in the MS and in the RNC.

Authentication functions

No execution of cryptographic functions nor storage of long-term secret keys is needed in the SN. This means that there is no need to standardise an authentication algorithm, and every operator is free to choose his own. However, for the reasons explained in Section 2, an example algorithm was included in the standards.

Compatibility with GSM security architecture

Especially in the initial years of UMTS, coverage will be provided only in isolated 'islands' so that handover

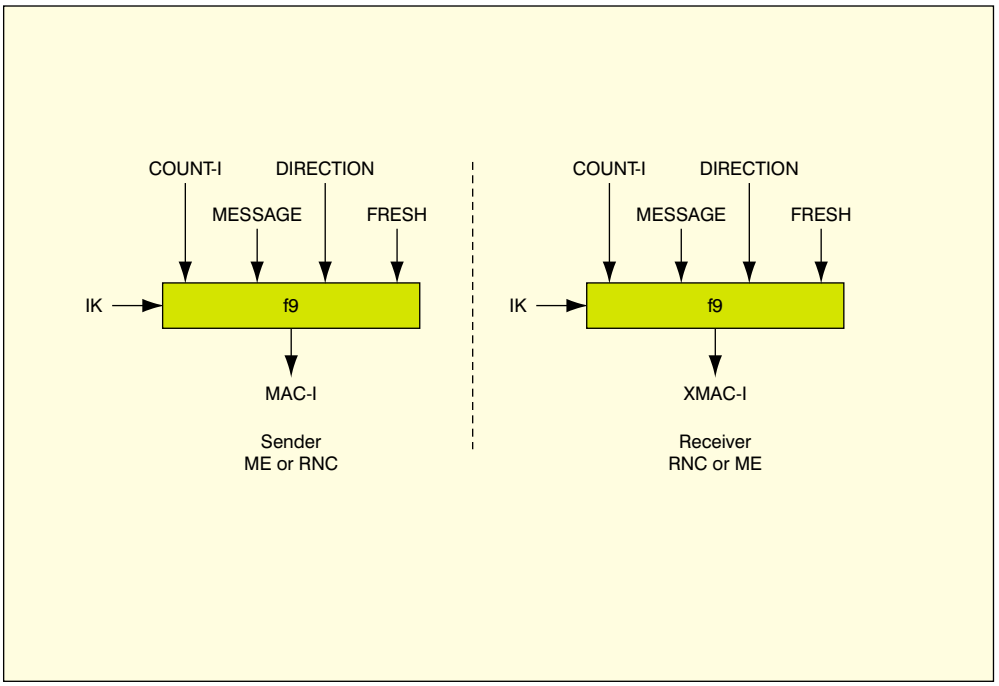

Fig. 2 Integrity protection of signalling traffic

between UMTS and GSM will be needed frequently. The UMTS AKA has been designed in such a way that roaming and handover between GSM and UMTS works as smoothly as possible as far as security is concerned. This is facilitated by the similarity between UMTS quintets and GSM triplets. Conversion functions are specified to convert quintets into triplets and vice versa. (In the case of the conversion of a triplet into a quintet, of course, only GSM-grade security can be achieved.)

Compatibility with 3GPP2

As well as supporting roaming with second generation GSM systems, it is also desirable for UMTS systems to support roaming to the 3G system being developed by 3GPP2. Rather than develop a different security architecture, 3GPP2 has adopted the 3GPP authentication and key agreement protocol as the basis for its security architecture to facilitate roaming between the two systems.

Integrity protection in the access network

Requirements for integrity protection

Integrity protection in UMTS prevents the insertion, modification, deletion and replay of signalling messages exchanged between the MS and the RNC. The reasons for supplementing the existing ciphering mechanism with a dedicated integrity mechanism in UMTS are summarised below:

- For various reasons UMTS networks must be able to instruct the MS to use an unciphered connection, i.e. the use of ciphering cannot be made mandatory. Thus, an active man-in-the-middle attacker could potentially compromise user traffic confidentiality by masquerading as a network to establish an unciphered connection towards the user. Since integrity protection can be made mandatory, this attack can be prevented as the user can always verify the instruction from the network to establish an unciphered connection. In GSM the instruction from the network to establish an unciphered connection is not integrity protected.
- The ability to integrity-protect ciphering algorithm negotiation messages provides protection against bidding-down attacks where an active attacker forces the use of an old ciphering algorithm, which may, for instance, allow user traffic confidentiality to be compromised. This feature only becomes of interest when multiple algorithms are supported in the system, as is the case in GSM. In the first release of the 3GPP standards only one ciphering algorithm is available and all mobile stations must support this. However, it was considered desirable to design a future-proof system, which allowed new algorithms to be deployed in a way that protects against bidding down attacks.
- Although ciphering of signalling traffic provides some integrity protection and the ciphering of user traffic severely limits the usefulness of any successful compromise of signalling message integrity, the application of a dedicated integrity protection mechanism with its own integrity key increases the security margin of the system. This is seen as an important enhancement, which will ensure that 3G offers adequate protection against increasingly sophisticated active attackers.
- Although the application of user traffic ciphering is highly recommended not just for confidentiality but also for authentication and integrity purposes, there may be some exceptional cases where it is not applied. In these cases, integrity protection of signalling messages significantly increases the level of resistance against relatively unsophisticated attacks that would have been effective had integrity protection not been provided.

With the above requirements in mind, the mechanisms that are provided for integrity protection by the 3GPP system will be reviewed.

Integrity protection of signalling traffic

Once an integrity key has been established as part of an authentication protocol run and once the available integrity protection algorithms in the MS are known, the network can start integrity protection. Integrity protection is applied in the mobile equipment (ME) at the user side and in the radio network controller (RNC) at the network side. A message authentication code function is applied to each individual signalling message at the radio resource control (RRC) layer in the UTRAN (UMTS Terrestrial Radio Access Network) protocol stack[15,16].

After the RRC connection establishment and the execution of the security mode establishment procedure, most of the subsequent RRC signalling messages are integrity protected. This includes the RRC signalling messages themselves, plus so-called RRC direct transfer messages, which contain protocol data units for higher layer dedicated signalling between the ME and the core network. Protection of direct transfer messages allows mobility management, call control and session management signalling to be protected.

Fig. 2 illustrates the use of the integrity algorithm f9 to authenticate the data integrity of an RRC signalling message. The input parameters to the algorithm are:

- the integrity key IK, which is 128 bits long
- an integrity sequence number (COUNT-I) and a random value generated by the radio network controller (FRESH). COUNT-I and FRESH are each 32 bits long. Together, they provide replay protection.
- a direction identifier (DIRECTION) to prevent so-called reflection attacks
- the RRC signalling message content (MESSAGE).

Based on these input parameters the sender computes the 32 bit message authentication code for data integrity (MAC-I) using the integrity algorithm f9. The MAC-I is then appended to the RRC message when sent over the radio access link. The receiver computes the expected MAC-I (XMAC-I) on the message received in the same way as the sender computed MAC-I on the message sent and verifies the data integrity of the message by comparing it to the received MAC-I.

The integrity feature also provides data origin authentication, so that the receiver of an integrity-protected signalling message can corroborate the identity of the sender. This allows an operator not to run the full authentication and key agreement protocol every time a connection is established.

Partial integrity protection of user traffic

UMTS also has a mechanism which prevents the insertion or deletion, but not the modification, of user traffic. This feature is meant to prevent certain bandwidth hijacking attacks while avoiding the cost of full-blown integrity protection mechanisms for user data.

The procedure works by allowing the RNC to monitor the sequence numbers for integrity protection and ciphering associated with each radio bearer. The RNC may send an integrity-protected signalling message to the ME. The message contains the most significant parts of the counter values, which reflect the amount of data sent and received on each active radio bearer. On reception the ME checks that the counter values agree with the values maintained by the ME. If there is a difference then this is indicated in an appropriate response.

Ciphering in the access network

Ciphering, when applied, is performed in the RNC and the ME. The ciphering function is performed either in the radio link control (RLC) sublayer (for non-transparent RLC mode) or in the medium access control (MAC) sublayer (for transparent RLC mode)[16].

Fig. 3 illustrates the use of the ciphering algorithm f8 to encrypt plaintext by applying a key stream using a bit-per-

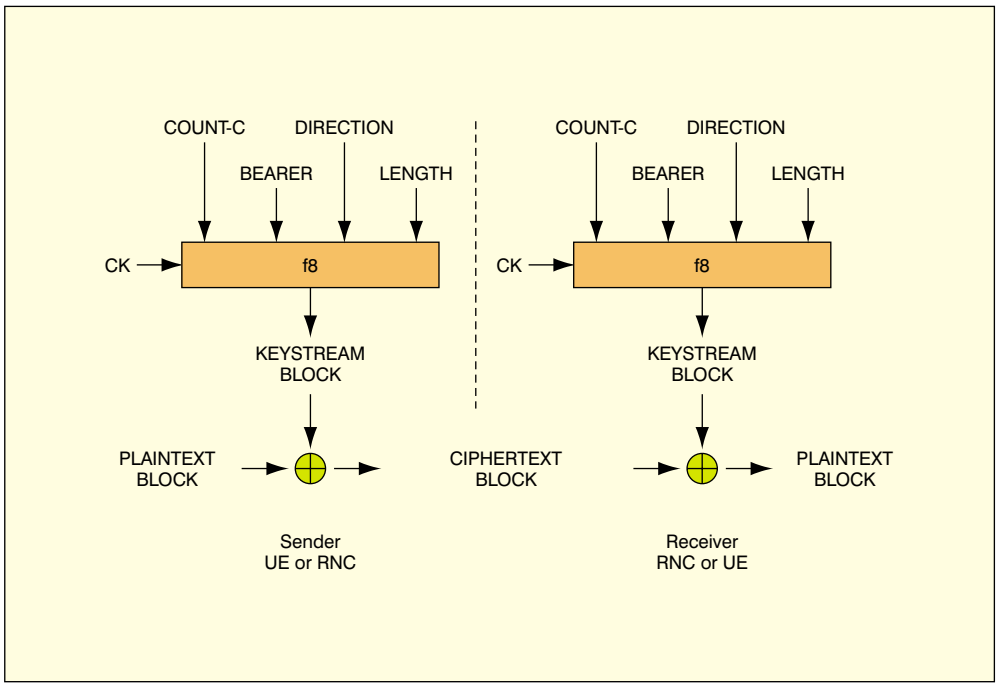

Fig. 3 Ciphering of user and signalling traffic

bit binary addition of the plaintext and the key stream. The plaintext may be recovered by generating the same key stream using the same input parameters and applying a bit-per-bit binary addition with the ciphertext.

The input parameters to the algorithm are:

- the cipher key CK, which is 128 bits long
- the time-dependent input COUNT-C of length 32 bits
- the bearer identity BEARER
- the direction of transmission DIRECTION; and
- the length of the required key stream LENGTH.

Based on these input parameters the algorithm generates the output key stream block KEYSTREAM, which is used to encrypt the input plaintext block PLAINTEXT to produce the output ciphertext block CIPHERTEXT.

The input parameter LENGTH affects only the length of the KEYSTREAM BLOCK, not the actual bits in it.

Confidentiality and integrity algorithms

A common algorithm called KASUMI forms the basis for both the confidentiality algorithm f8 and the integrity algorithm f9, which are standardised for use in 3GPP systems. KASUMI is a block cipher[17], which is used in two different operating modes to construct f8 and f9[18].

4 Network domain security

The term 'network domain security' in the 3GPP specifications covers security of the communication between network elements. In particular, the mobile station is not affected by network domain security. The two communicating network elements may both be in the same network administrated by a single mobile operator or they may belong to two different networks. The latter case, i.e. internetwork communication, clearly requires standardised solutions, because otherwise each pair of two operators that are roaming partners would need to agree on a common solution. The intranetwork case also benefits from standardisation as many operators have network elements manufactured by several different vendors.

In the past there have been no cryptographic security mechanisms available for internetwork communication. The security has been based on the fact that the so-called SS7 network has been accessible only to a relatively small number of well-established institutions. It has been very difficult for an attacker to insert or manipulate SS7 messages. The situation is changing now for two reasons. First, the number of different operators and service providers that need to communicate with each other is increasing. Second, there is a trend to replace Signalling System Number 7 (SS7) networks with Internet Protocol (IP) networks. The introduction of IP brings many benefits but it also means that a large number of hacking tools, some of which are available on the Internet, become applicable to the telecommunication networks. For instance, various denial of service attacks may be anticipated. For these reasons, the lack of cryptographic protection for internetwork signalling may increasingly become a security risk.

A major part of the 3GPP Release 99 specifications was devoted to the introduction of a completely new radio access technology, while the core network part was an extension of the existing GSM specification set. This is the main reason why the protection mechanisms for core network signalling were not introduced in Release 99 but instead in later releases, starting from Release 4.

The mobile specific part of SS7 signalling is called the mobile application part (MAP). In order to protect all communication in SS7 networks it is clearly not enough to protect only the MAP protocol. However, from the point of view of mobile communications, MAP is the essential part to be protected. For instance, the session keys for protecting the radio interface and other authentication data are carried in MAP. On the other hand, the specification of a security protocol for SS7 would have been a big task unlikely to be completed in the required time-frame. Mainly for these reasons, 3GPP has developed security mechanisms that are specific to MAP. The functional description of these mechanisms (stage 2) is given in the TS 33.200[19] while the bit-level materialisation (stage 3) is described in the MAP specification itself TS 29.002[20]. The whole feature is called MAPSEC and the first release in which it is included in 3GPP is Release 4. Note that the MAPSEC protocol protects MAP messages at the application layer. An overview of MAPSEC is given in the next subsection.

Many different security mechanisms have been standardised by the Internet Engineering Task Force (IETF) for IP-based networks. Hence, there is no need to specify a completely new solution for 3GPP. On the other hand, it is still important to agree on how IETF protocols are used to protect IP-based communication in 3GPP networks. Specification TS 33.210[21] is devoted to this task. It is included in the 3GPP Release 5 specification set. The main tool from the IETF used in 3GPP is the IPsec protocol suite[22].

Note that 3GPP also specifies how the MAP protocol can be run on top of IP. In this case, there are basically two alternative methods to protect MAP: either to use MAPSEC or IPsec. The latter has the advantage that the protection also covers lower layer headers as is done in the IP layer.

MAPSEC

The basic idea of MAPSEC can be described as follows. The plaintext MAP message is encrypted and the result is put into a 'container' in another MAP message. At the same time a cryptographic checksum, i.e. a message authentication code covering the original message, is included in the new MAP message. To be able to use encryption and message authentication codes, keys are needed. MAPSEC has borrowed the notion of a security association (SA) from IPsec. The SA contains cryptographic keys, but in addition it contains other relevant information such as key lifetimes and algorithm identifiers. Security associations of MAPSEC resemble IPsec SAs but the two are not identical.

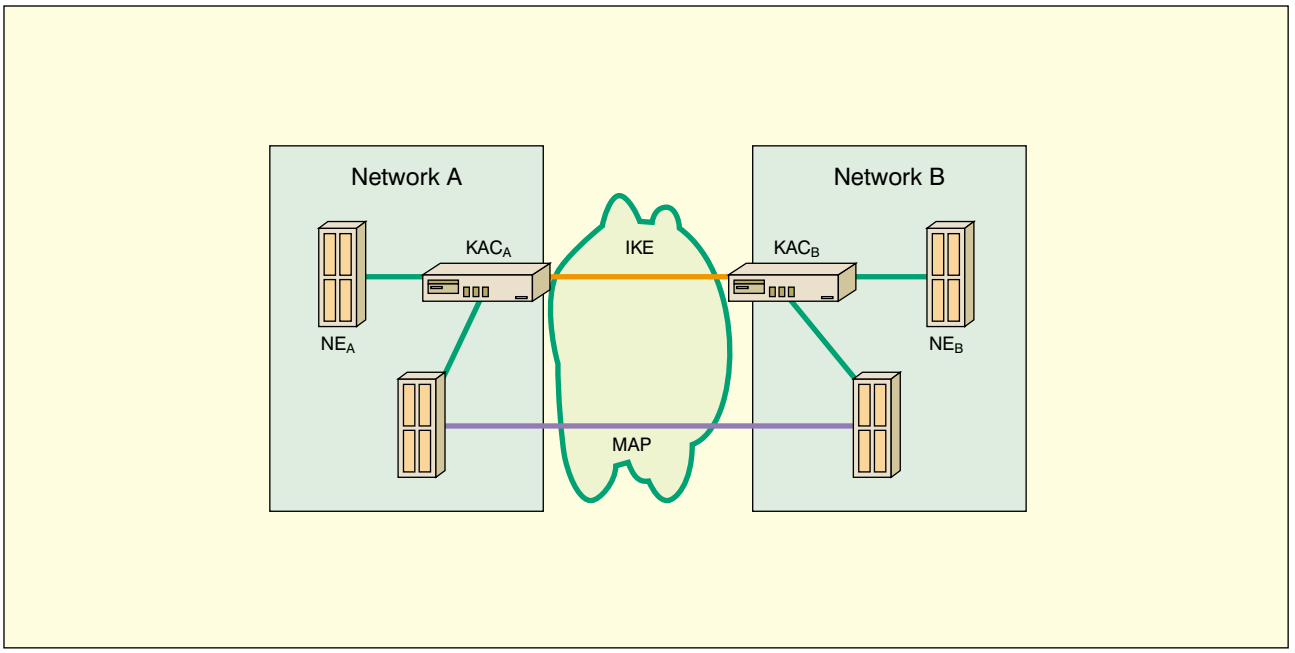

Fig. 4 Automatic key management for MAPSEC

3GPP Release 4 does not specify how SAs are to be exchanged between operators. In practice, this implies that the SAs are configured in the network elements manually. The automatic key management for MAPSEC (Fig. 4) will be fully specified in the 3GPP Release 6 specification set. The basic ingredient in MAPSEC automatic key management is a new element called a key administration centre (KAC). These KACs agree on SAs between themselves using the IETF Internet key exchange (IKE) protocol[23]. The KACs also distribute the SAs to the network elements. All elements in the same security domain, e.g. elements in one operator's network, share the same SAs; they also share the policies on how to handle these SAs and incoming messages. The sharing of SAs is unavoidable as only networks, not individual network elements, can be addressed in MAP messages.

MAPSEC has three protection modes: no protection, integrity protection only and encryption with integrity protection. MAP messages in the last mode have the following structure: Security header || f6(Plaintext) || f7(Security header || f6(Plaintext)), where f6 is the Advanced Encryption Standard (AES) algorithm in counter mode and f7 is AES in CBC-MAC mode. The security header contains information needed to be able to process the message at the receiving end, such as the security parameters index, the sending network element identifier and the time variant parameter.

In MAPSEC only some of the MAP operations are protected, including the most critical operations, such as the authentication data transfer. This is for performance reasons.

IPsec-based mechanisms

In the IPsec-based solution, all control plane IP communication towards external networks should go via a new element called a security gateway (SEG) (Fig. 5). These gateways use the IKE protocol[23] to exchange IPsec SAs between themselves. An important conceptual distinction between a security gateway and a MAPSEC KAC is that the former uses the negotiated SAs while the latter distributes the negotiated SAs to other elements, which send and receive the actual MAPSEC messages. In 3GPP Release 5 the IKE is based on preshared secrets but in future releases support of a public key infrastructure (PKI) for key management may be added.

A security gateway contains both an SA database and also a security policy database, which indicates how and when the SAs are used or have to be used. Naturally, the security gateway has to be physically secured. Typically the security gateway would be combined with firewall functionality.

One obstacle in reaching full interoperability of IPsec is the great number of options in the specifications. In 3GPP cutting down the number of options has solved this problem. The basic selections can be summarised as follows:

- Only Encapsulating Security Payload (ESP)[24] is used for protection of packets, while Authentication Header (AH)[25] is not used at all.
- ESP is always used in tunnel mode.
- AES is chosen as the encryption algorithm.
- IKE is used for key exchange in main mode phase 1 with preshared secrets.

It should be noted, however, that operators might configure more options (e.g. transport mode) in their own networks. The specification TS 33.210 describes only the core part that guarantees interoperability between different security domains.

5 IP multimedia subsystem security

The use of SIP for multimedia session control in 3GPP

The IP multimedia subsystem (IMS) is a core network

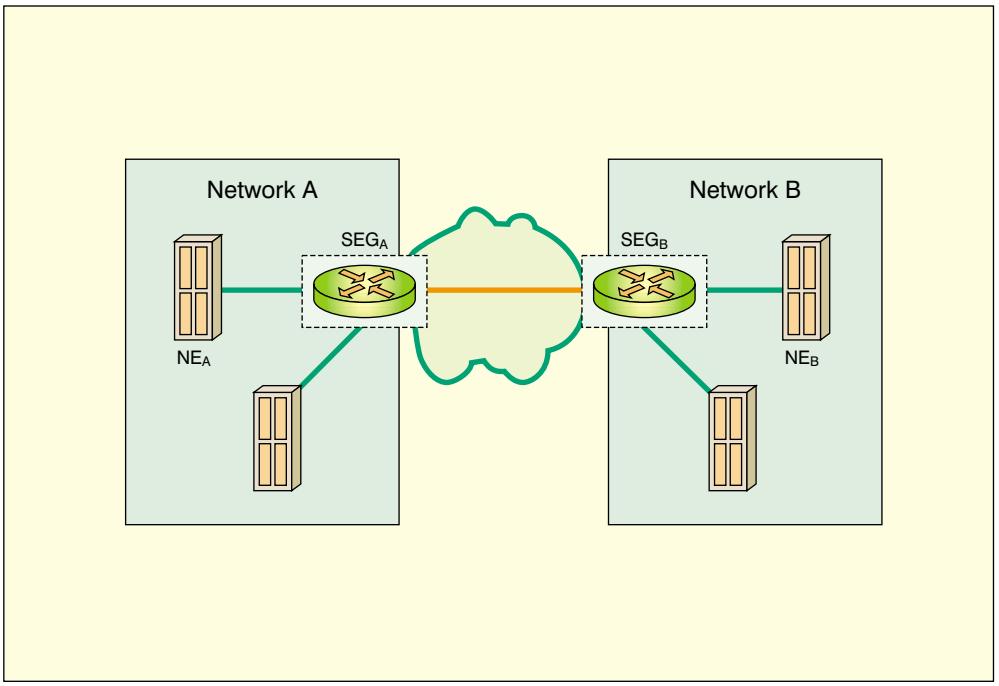

Fig. 5 Network domain security for IP-based control messages

subsystem within UMTS. It is based on the use of the Session Initiation Protocol (SIP)[26] to initiate, terminate and modify multimedia sessions such as voice calls, video conferences, streaming and chat. SIP is specified by the Internet Engineering Task Force (IETF)[27]. IMS also uses the IETF Session Description Protocol (SDP)[28] to specify the session parameters and to negotiate the codecs to be used. SIP runs on top of different IP transport protocols such as the User Datagram Protocol (UDP) and the Transmission Control Protocol (TCP).

The IMS architecture is specified in TS 23.228[29] and is built upon the UMTS packet domain. However, the architecture of the IMS is designed such that in future releases it may use mechanisms for IP connectivity other than those provided by the UMTS packet domain. This requirement is known as 'access network independence'. In order to meet this requirement the design relies to a large extent on IETF mechanisms. The 3GPP community has been very actively involved in work at the IETF to provide IETF standards that meet 3GPP requirements.

SIP itself is based on an IETF architecture that is very general and from which several trust models may be defined leading to hop-by-hop, end-to-middle and end-to-end security solutions. The IETF SIP working group has therefore defined several security mechanisms that can be applied to the different use cases of SIP. The mechanisms offer, for example, authentication, confidentiality and integrity of messages and replay protection. It should be stressed that SIP requests and responses cannot be fully encrypted or integrity protected on an end-to-end basis since parts of the messages by definition have to be available to proxies for routing purposes and for modification of messages. At the application layer it is possible to use HTTP (HyperText Transfer Protocol) authentication and S/MIME (since SIP carries MIME bodies). S/MIME (Secure/Multipurpose Internet Mail Extensions) has the disadvantage that it is based on public key certificates and may in certain cases generate very large messages (which has to be avoided over a wireless channel). At the lower transport and network layers, it is theoretically possible to use either TLS (Transport Layer Security) or IPsec to secure the whole SIP message. Although both UDP and TCP may be used in IMS, UDP is the default protocol and TCP is only used for large messages. As TLS can only be used on top of TCP, it has to be ruled out. TLS can also be ruled out because it uses public key certificates.

A 3GPP IMS subscriber has one IP multimedia private identity (IMPI) and at least one IP multimedia public identity (IMPU). To participate in multimedia sessions, an IMS subscriber must register at least one IMPU with the IMS. The private identity is used only for authentication purposes.

There are four IMS entities relevant to the IMS security architecture:

- UE: The user equipment (UE) contains the SIP user agent (UA) and the smart card based IMS subscriber identity module (ISIM), an application that contains the IMS security information. The ISIM can be a distinct application sharing no data and functions with the USIM, or it can share data and security functions with the USIM or it can be a reused USIM. There can only be one ISIM per IMPI.
- P-CSCF: The proxy call session control function (P-CSCF) acts as an outbound SIP proxy. For the UA in the UE, it is the first contact point in the serving network. It forwards SIP requests towards the I-CSCF.
- I-CSCF: The interrogating call session control function (I-CSCF) is the contact point in the home network and acts as a SIP proxy. It forwards SIP requests or responses towards a S-CSCF.
- S-CSCF: The serving call session control function (S-CSCF) may behave as a SIP registrar, a SIP proxy server and a SIP UA. Before the UE can send a SIP INVITE message to set up a session it has first to

register an IMPU at the S-CSCF. The registration of an IMPU is done by the UE by sending a SIP REGISTER message towards the home network.

Sessions are set up using INVITE messages. Fig. 6 describes a scenario where an INVITE is sent from one user equipment to another, both of which reside in a 3GPP network. The INVITE from UE A in Home A first passes through a P-CSCF and then to an I-CSCF, which forwards the message to the home subscriber system (HSS), which looks up to which S-CSCF the user is registered. A similar process is done in Home B and the INVITE is finally terminated in UE B. The conversation may now start, for example, by using the IETF Realtime Transport Protocol (RTP)[30].

Security architecture for the IP multimedia subsystem

The 3GPP IMS security architecture is specified in TS 33.203[31]. An IMS subscriber will have a private identity (IMPI), which is authenticated. All relevant subscriber data is stored in the HSS. During user registration, which will take place in the S-CSCF, the subscriber data is transferred from the HSS to the S-CSCF. Hence upon request by a user the S-CSCF can match this request with the subscriber profile before access is granted, such that the home network can control access.

The registration process is shown in Fig. 7. In SM1 the UE sends an unprotected REGISTER, which is forwarded towards the S-CSCF. The authentication and key agreement for IMS is based on the same mechanism as in UMTS, i.e. the AKA protocol described in Section 3. The challenge is derived in the HSS when the UE REGISTERs for the first time. This challenge is forwarded by the S-CSCF towards the P-CSCF and the UE (SM4—SM6). The UE then checks that the challenge is authentic and sends the response back to the S-CSCF (SM7), who will authenticate the subscriber. This is somewhat different to the UMTS architecture where the authentication is delegated to the serving network. Hence the trust in the P-CSCF is somewhat reduced. The S-CSCF in the home network can at any time require that the user is re-authenticated.

When the user has been successfully authenticated and a 200 OK has been received by the UE a security association (SA) is active between the UE and the P-CSCF for the protection of subsequent SIP messages between these two entities. The P-CSCF will obtain the integrity key for protecting SIP messages from the SM5 message sent by the S-CSCF towards the UE. Since the integrity key is passed from the S-CSCF to the P-CSCF this message has to be protected between the home network and the serving network. This is accomplished by using

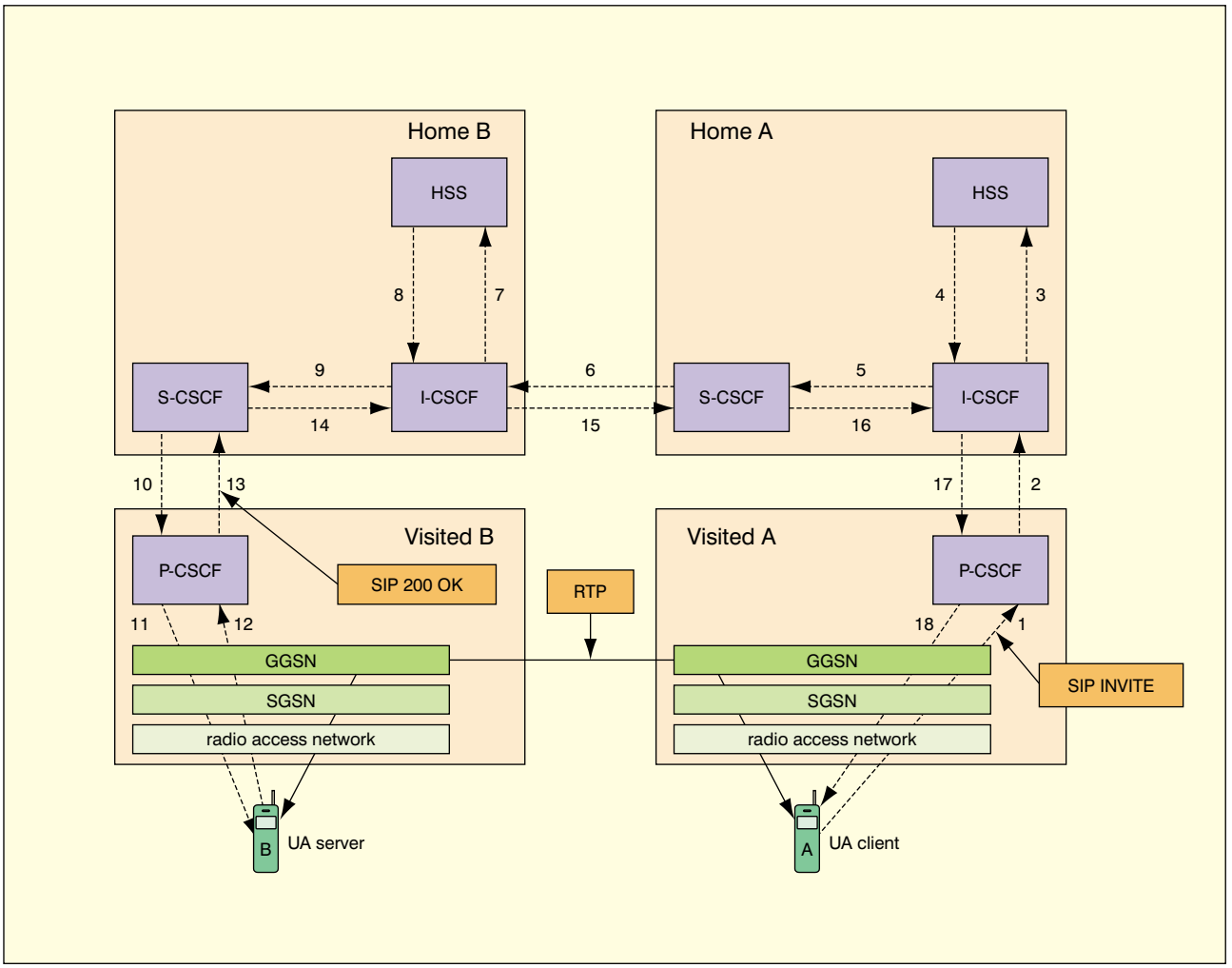

Fig. 6 Session establishment in IMS

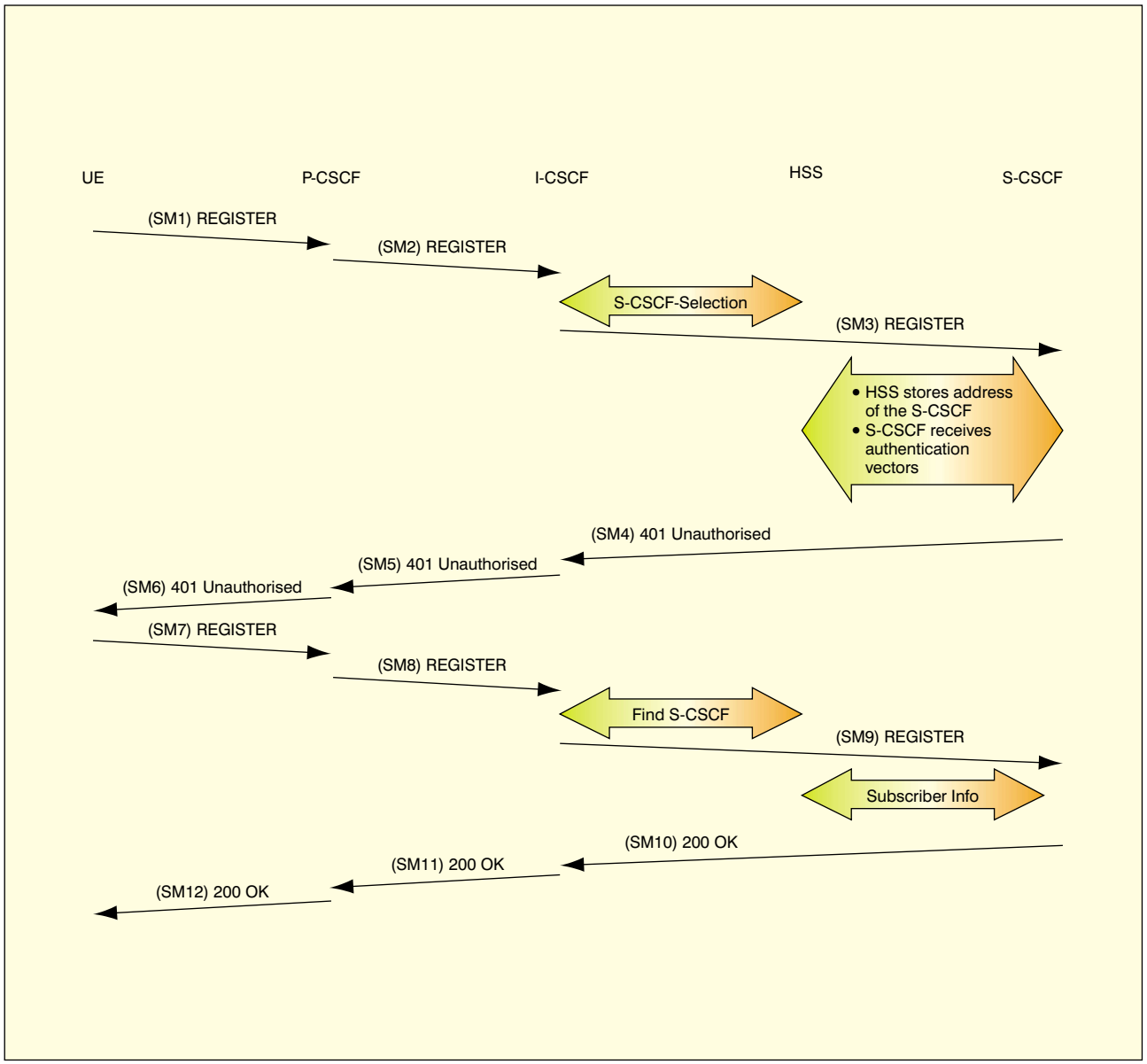

Fig. 7 Registration process in IMS

IPsec tunnels between the networks using the mechanisms specified in TS 33.210[21] and described in Section 4 of this paper.

Only integrity protection is used between the UE and the P-CSCF and this feature is important from a billing and charging perspective. No confidentiality protection is offered in IMS. Instead, an operator should use the confidentiality mechanism offered in UMTS at the link layer that is terminated in the RNC. For end-to-end security the IETF specifications SRTP[32] and MIKEY[33] can secure RTP and provide an appropriate key management mechanism.

The protocol that is used for providing integrity protection of the SIP application control plane in the IMS is IPsec ESP[24]. IPsec ESP requires security associations, which can be created using either IKE (Internet Key Exchange)[23] or 'manual keying'. An IKE implementation is quite complex and is therefore not feasible for mobile terminals. As a consequence, 'manual keying' is used in IMS. The term 'manual keying' is IPsec terminology and may be misleading here: in fact, key management is not manual, but automatically provided by the AKA protocol as described below. A local application programming interface (API, not specified in 3GPP) is used to transport the key and other information that is required by IPsec from the SIP layer to the IPsec layer.

Authentication using HTTP Digest AKA

SIP is based on the HTTP framework and therefore SIP inherited the authentication framework of HTTP[34]. However, the IETF SIP working group has forbidden the use of HTTP Basic authentication because it is an insecure protocol due to the fact that the password is sent in clear. Hence only HTTP Digest authentication is allowed within SIP. It should be noted that AKA is perfectly secure even though the parameter RES is sent in clear and hence the use of HTTP Basic together with 3GPP AKA would not cause any security weaknesses. However, 3GPP has aimed to be compliant with IETF specifications as far as possible and therefore HTTP Basic is not used.

The use of HTTP Digest together with 3GPP AKA[35] is

described in the following. HTTP Digest is used for sending the authentication challenge from the S-CSCF towards the UE in a 401 Unauthorised response associated with the received REGISTER request. This message includes the WWW-Authenticate header, which transports, for example, the authentication challenge, the security mechanism identifier (which is called AKAv1-MD5) and the integrity key (IK) for the P-CSCF. Upon receiving the authentication challenge the UE first checks that the sequence number is in the correct range and that the message authentication code is correct. If these checks are successful the UE derives the authentication response RES based on the challenge and the long-term authentication key stored in the ISIM. The RES is treated as a password in the HTTP Digest framework and a response is calculated by using AKAv1-MD5 with the RES as the input. The S-CSCF has to make similar calculations based on the expected RES (XRES) before a check can be performed.

Integrity protection using IPsec ESP

IPsec Encapsulated Security Protocol (ESP)[24] is applied in IMS in transport mode between the UE and the P-CSCF. A pair of unidirectional IPsec security associations (SAs) between the UE and the P-CSCF is simultaneously established during authentication. The integrity key is the same for both security associations. In order to resist replay attacks the IPsec ESP anti-replay service is used. Furthermore since the keys are the same in the two SAs, protection against reflection attacks is achieved by ensuring that the security parameters index (SPI) is different for each SA. It is specifically forbidden to use the NULL authentication algorithm since the SIP signalling originating from a user is the basis for charging and billing.

The authentication (i.e. integrity) algorithms that are used are either HMAC-MD5-96[36] or HMAC-SHA-1-96[37].

The SA is bound to the normal IPsec selectors, i.e. source and destination IP addresses, and source and destination ports. The allowed transport protocols in IMS are UDP and TCP.

Not all messages can be integrity protected so the security architecture has to allow some unprotected messages to proceed (e.g. error messages). The P-CSCF shall accept unprotected messages on a port different from the port used for protected messages. A similar approach is applicable for the UE.

Security mode set-up and use of SIP security agreement

The mechanism for setting up security associations (SAs) for IPsec ESP in the IMS is based on the use of three new SIP headers to negotiate different security mechanisms in a secure manner[38]. The mechanism defined for use within 3GPP is manually keyed IPsec without IKE, which is known as 'ipsec-man'. The negotiation procedure allows new mechanisms to be introduced in a way that is resistant to bidding down attacks.

6 Further developments in UMTS security

This paper surveys the major security features that are included in the first releases of the UMTS standards.

Work on the next UMTS release has recently started. This will introduce new security features. Many of these features will be introduced to secure the new services that will be introduced, e.g. presence services, push services and multicast/broadcast services.

Looking more into the future, mobile cellular systems will have to accommodate a variety of different radio access networks, including short-range wireless technologies, connected to a common core network. On the user side the concept of a monolithic terminal, as we know it, is dissolving. Distributed terminal architectures are appearing whose components are interconnected by short-range radio links. These new developments represent a major challenge to the UMTS security architecture. A collaborative research project funded by the European Union and called SHAMAN (Security for Heterogeneous Access in Mobile Applications and Networks)[39] is currently tackling these issues.

A separate project is also underway to identify research topics in the area of mobile communications to be conducted as part of the European Union's 6th Framework Programme of sponsored collaborative research; the project is called PAMPAS (Pioneering Advanced Mobile Privacy and Security)[40].

References

1 http://www.itu-2000.org
2 RICHARDSON, K. W.: 'UMTS overview', *Electron. Commun, Eng. J.*, June 2000, **12,** (3), pp.93–100
3 http://www.etsi.org
4 http://www.gsm.org
5 http://www.3gpp.org
6 http://www.3gpp2.org
7 GSM 03.20: 'Network related security features'. See http://www.3gpp.org
8 PIPER, F., and WALKER, M.: 'Cryptographic solutions for voice telephony and GSM'. Proc. COMPSEC'98 (Elsevier, 1998)
9 WALKER, M., and WRIGHT, T.: 'Security aspects', *in* HILLEBRAND, F. (Ed.): 'GSM and UMTS: The creation of global mobile communication' (John Wiley & Sons, 2002)
10 3GPP TS 35.205-35.208: 'Specification of the MILENAGE algorithm set; an example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5 and f5*'. See http://www.3gpp.org
11 3GPP TS 33.102: 'Security architecture'. See http://www.3gpp.org
12 HORN, G., and HOWARD, P.: 'Review of third generation mobile system security architecture'. Proc. ISSE'00, Barcelona, 27th–29th September 2000
13 GOLLMANN, D.: 'What do we mean by entity authentication'. IEEE Symposium on Security and Privacy, Oakland, CA, 6h–8th May 1996
14 MENEZES, A., VAN OORSCHOT, P., and VANSTONE, S.: 'Handbook of applied cryptography' (CRC Press, Boca Raton, 1997)
15 3GPP TS 25.301: 'Radio Interface Protocol Architecture'. See http://www.3gpp.org
16 3GPP TS 25.331: 'RRC Protocol specification'. See http://www.3gpp.org
17 3GPP TS 35.202: 'Specification of confidentiality and integrity algorithms: KASUMI specifications'. Available under licence, see http://www.3gpp.org

Krister Boman received an MSc degree in Mechanical Engineering (branch of specialisation Mathematics) from Chalmers University of Technology, Gothenburg, Sweden, in 1992 and a Licentiate of Engineering degree from Chalmers University of Technology, Department for Mathematics in 1995. He joined Ericsson in 1995 where his work has mainly related to signal processing and cryptography, including managerial positions. Recently he has taken an active rôle in the development of security solutions for 3G networks and currently represents Ericsson in the 3GPP security group.

Address: Ericsson AB, SE-412 66 Gothenburg, Sweden
E-mail: krister.boman@erv.ericsson.se

Valtteri Niemi received an MSc degree from the University of Turku, Mathematics Department, in 1987 and a PhD degree from the same department in 1989. He held several research and teaching posts at the University of Turku, including acting as an Associate Professor in Mathematics for the academic year 1992–3. In 1993 he was nominated as an Associate Professor in the Mathematics and Statistics Department of the University of Vaasa, where he stayed until joining Nokia Research Center, Mobile Networks laboratory, Helsinki in 1997. He now works on security issues in future mobile networks and terminals, the main emphasis being on cryptological aspects. He is a vice chairman of the 3GPP TSG SA WG3 (security) standardisation group. In addition to cryptology and security, Dr. Niemi has done research in the area of formal languages.

Address: Nokia Research Center, PO Box 407, FIN-00045, Helsinki, Finland
E-mail: valtteri.niemi@nokia.com

Peter Howard graduated from Aston University, Birmingham in 1995 with an MEng degree in Electronic Systems Engineering. Since joining Vodafone in 1995 he has worked in several collaborative research projects on mobile security sponsored by the European Union, including ASPeCT, USECA and SHAMAN. He is also active in the standardisation of UMTS security in 3GPP TSG SA WG3. He is an IEE Member and a Chartered Engineer.

Address: Vodafone Group Research and Development, The Courtyard, 2–4 London Road, Newbury, Berkshire, RG14 1JX, UK
E-mail: peter.howard@vodafone.com

Günther Horn studied Mathematics and Physics at the universities of Stuttgart, London and Tübingen. He received a PhD (Dr.rer.nat.) degree in Mathematics from the University of Tübingen in 1984. He was a Visiting Assistant Professor of Mathematics at the University of California at Irvine in 1985. He joined the Corporate Technology Labs of Siemens AG in Munich in 1986, since when he has been engaged in research on new telecommunication systems. For the past eight years, Dr. Horn has worked mainly on security in mobile networks, in particular on authentication protocols for UMTS, fraud control, electronic payments for mobile applications and security issues related to the use of GSM for future train control systems. He has led work packages in several collaborative research projects sponsored by the European Union, including ASPeCT, USECA and SHAMAN. He is also active in the standardisation of UMTS security in 3GPP TSG SA WG3.

Address: Siemens AG, Corporate Technology, D-81730 München, Germany
E-mail: guenther.horn@siemens.com

18 3GPP TS 35.201: 'Specification of confidentiality and integrity algorithms: f8 and f9 specifications'. Available under licence, see http://www.3gpp.org
19 3GPP TS 33.200: 'Network domain security: MAP application layer security'. See http://www.3gpp.org
20 3GPP TS 29.002: 'Mobile Application Part (MAP) specification'. See http://www.3gpp.org
21 3GPP TS 33.210: 'Network domain security: IP network layer security'. See http://www.3gpp.org
22 IETF RFC 2401-2412 (1998): 'IPsec protocol suite'. See http://www.ietf.org
23 IETF RFC 2409 (1998): 'IKE: Internet Key Exchange'. See http://www.ietf.org
24 IETF RFC 2406 (1998): 'IP Encapsulating Security Payload (ESP)'. See http://www.ietf.org
25 IETF RFC 2402 (1998): 'IP Authentication Header'. See http://www.ietf.org
26 CAMARILLO, G.: 'SIP demystified' (McGraw-Hill, 2002)
27 IETF RFC 3261 (2002): 'SIP: Session Initiation Protocol'. See http://www.ietf.org
28 IETF RFC 2327 (1998): 'SDP: Session Description Protocol'. See http://www.ietf.org
29 3GPP TS 23.228: 'IP Multimedia Subsystem (IMS), Stage 2'. See http://www.3gpp.org
30 IETF RFC 1889 (1996): 'RTP: a transport protocol for real-time applications'. See http://www.ietf.org
31 3GPP TS 33.203: 'Access security for IP-based services' See http://www.3gpp.org
32 IETF draft-ietf-avt-srtp-04 (2002): 'The Secure Real Time Transport Protocol'. See http://www.ietf.org
33 IETF draft-ietf-msec-mikey-01 (2002): 'MIKEY: Multimedia Internet KEYing'. See http://www.ietf.org
34 IETF RFC 2617 (1999): 'HTTP authentication: Basic and Digest access authentication'. See http://www.ietf.org
35 IETF draft-ietf-sip-digest-aka-03 (2002): 'HTTP Digest authentication using AKA'. See http://www.ietf.org
36 IETF RFC 2403 (1998): 'The use of HMAC-MD5-96 within ESP and AH'. See http://www.ietf.org
37 IETF RFC 2404 (1998): 'The use of HMAC-SHA-1-96 within ESP and AH'. See http://www.ietf.org
38 IETF draft-ietf-sip-sec-agree-04 (2002): 'Security mechanism agreement for SIP sessions'. See http://www.ietf.org
39 http://www.ist-shaman.org
40 http://www.pampas.eu.org

Received 18th July 2002