Security for the Third Generation (3G) Mobile System

Colin Blanchard
Network Systems & Security Technologies
BTexaCT.
MLB1 PP8
Adastral Park
Ipswich
IP5 5RE
Phone +44 1473 605353
Fax +44 1473 623910
colin.blanchard@bt.com

1 Introduction

When considering security in mobile systems, in common with most other systems, our main objectives are in preventing:

- Access and use of service to avoid or reduce a legitimate charge.
- Loss of confidentiality or integrity of a user's or operator's data
- Denial of a specific user's access to their service or denial of access by all users to a service

However, user expectations for instant communication and ease of use, as well as terminals which are easily lost or stolen, present a number of unique challenges in the mobile environment.

The original first generation analogue mobile employed a simple electronic serial number to confirm that the terminal should be allowed access to the service. It was not long before the protection afforded to this number was broken. Eventually, devices appeared that could read these electronic serial numbers from the air, and access an unsuspecting user's account for a short time, before moving on to the next, in the hope that the small charges on each bill would not be noticed. So why was this not predicted at the time? Unfortunately, there always seems to be an assumption, with any new development in communications technology, that complexity alone will protect such services from abuse.

Second generation systems such as GSM were designed from the beginning with security in mind. This has stood up to the kind of attacks that were prevalent on the analogue system at the time, thanks mainly to the ability to put responsibility for security in the hands of the Home Environment (HE) operator. The HE operator can control the use of the system by the provision of the Subscriber Identity Module (SIM) which contains a user identity and authentication key. This is specifically arranged so that this long life authentication key is not required by the Serving Network (SN) when roaming, exposed over the air or exposed across the interface between the SIM and the mobile. This keeps to the minimum the level of trust the HE operator needs to place in the User, Serving Network and manufacturer of the Mobile Equipment (ME).

In 1996, when the 3rd Generation system known as UMTS was being developed in ETSI (European Telecommunications Standards Institute), the opportunity was taken to review the basis for security in existing mobile systems and to develop a new security architecture specifically to be used in UMTS. This early work was subsequently taken forward into the Third Generation Partnership Project (3GPP) and this will be the basis for the Release 99 deployment of 3G systems.

2 3G Security Principles

It was agreed that any new security architecture must be based on an evolution of GSM and must adopt four basic principles:

- It will take into account the additional features needed for actual or predicted change in the operating environment
- It will maintain compatibility with GSM wherever possible
- It will retain those features of GSM that have proved to be robust and useful to the user and network operator
- It will add or enhance features to overcome actual or perceived weaknesses in 2G system

2.1 Additional features

One of the main reasons for the development of the 3G system to make higher value services available to as many users as possible world wide, using a universal design of the handset. However, this increases the number of relationships, as the number of Users, Service Providers, and Network Operators in the market expands. This increased level of service interaction increases the number of potential attackers and the opportunities open to them. This was not too much of a problem in the initial roll out of GSM, as there were a relatively small number of operators and the risk of compromise was low. For 3G, the networks are getting smaller and more numerous, so opportunities for hackers and other abusers of networks will increase. Even if deliberate abuse is not considered likely, unintentional mishaps may occur as a result of the complexity and the rate of new service introduction.

2.2 Maintaining compatibility with GSM

A major contributor to the success of GSM has been the availability of a full system specification with standard service sets and automatic integrated roaming. An important consideration was to make as much use of the existing infrastructure as possible, while gradually enhancing the network as required meeting the demand for the new services. An example of this in GSM, has been the introduction of the General Packet Radio Service (GPRS) by the overlay of an IP core network and an additional Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) network elements. The existing radio system is virtually unchanged retaining the Home Location Registers /Visited Location Register (HLR/VLR) concept and operator control of security via SIM. This HLR/VLR concept provides rapid call set up and the retention of this was considered essential for evolution of 2G to 3G. The concept of authentication using a shared secret key was also retained. Consideration was given making use of the work done in the IP world, e.g. Public Key asymmetric techniques. However, the infrastructure was already in place to allow global roaming and recognition of security information. This was not the case with public key systems and a truly global Public Key Infrastructure (PKI) was seen to be some way off. Also, there were concerns whether the SIM card would be able to handle an asymmetric key protocol and still retain the performance the users are accustomed to. Figure 1 shows how these elements are retained to provide five specific aspects of security for operators and end users of the 3G network:

I. Network access security provides users with secure access to 3G services and protects against attacks on the (radio) access link;
II. Network domain security enables nodes in the provider domain to exchange signalling data securely;
III. User domain security allows users to have secure access to mobile stations;
IV. Application domain security enables applications in the user and in the provider domain to exchange messages securely;
V. Visibility and configurability of security informs the user if a security feature is in operation or not, allowing appropriate use of the service.

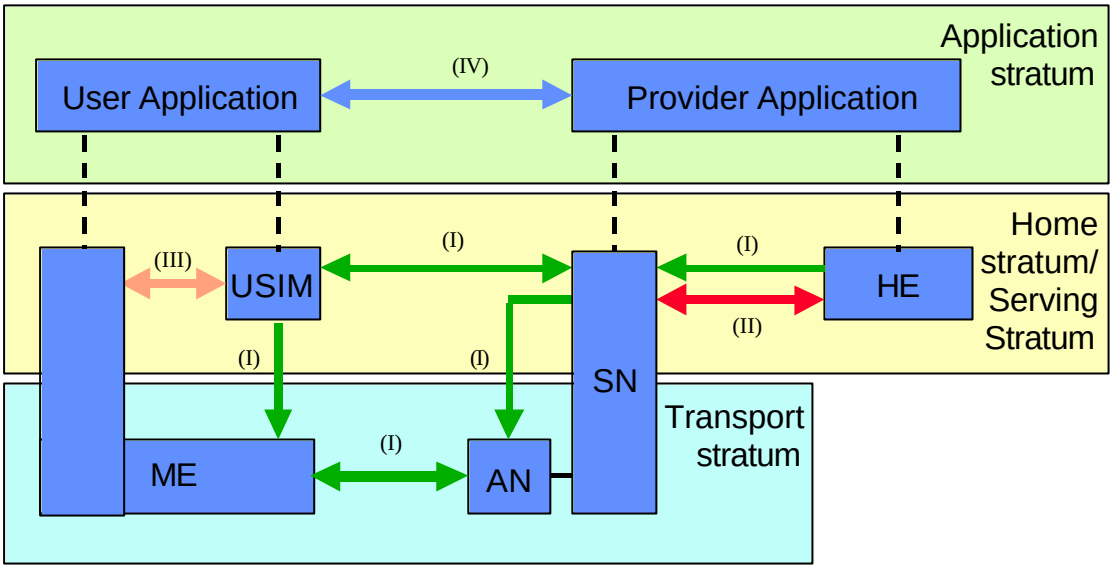

Figure 1: Overview of the security architecture

2.3 Retention of GSM features that have proved to be robust and useful

2.3.1 SIM based Authentication

One of the most important features is that of the SIM as a removable security module which is issued and managed by the Home Environment (HE) operator and is independent of the terminal. It was felt that this concept had been the most significant in maintaining the security of GSM, while retaining general user acceptance of the service. There is no need for any user action, other than perhaps entering an optional 4 digit PIN into the terminal. User guidance on security is no more than what they are familiar with from their bank: take care of your card, report its loss immediately and do not write the PIN down or disclose it to anyone.

The 3G system will retain the challenge and response authentication mechanism based on a symmetric secret key shared between the SIM and the Authentication Centre (AuC) in the Home Environment. The authentication key and algorithm for the challenge response mechanism are not required by the Serving Network which help keep the level of trust placed in the many possible serving networks to a minimum. This method also allows the algorithm to be made specific to the Home Environment, which means that the impact of any compromise can be confined to the user base of just one operator rather than the entire user community worldwide.

2.3.2 Confidentiality of user traffic on the air interface

Air interface encryption is being retained but not just for the benefit of the user. This confidentiality is actually essential for the network operator to be able to ensure that the validity of the authentication at the start of the call is maintained throughout the call i.e. to prevent a session from being hijacked. It proved impossible to reach an agreement to make this a mandatory feature, due to concern about restrictions on the use of encryption in some countries. The use of integrity protection on the signalling messages can be an alternative means of achieving this end.

2.3.3 Confidentiality of user identity on the air interface

Of course, any system that relies on looking up some security information e.g. decryption key or password in a database to decide if access is allowed requires that users identify themselves by an identity known to the system, just like a user id when accessing a computer system. In GSM, once this initial message is exchanged over the air, then a temporary user identity is allocated which is local to the area and is reassigned to another user as soon as the original user moves out of the area. This reduces the exposure of the real user identity on the air interface and prevents information on a user's use of a service or movements being gathered e.g. traffic flow analysis.

2.4 Addition and enhancement of features to overcome actual or perceived weaknesses in 2G system

Figure 2 shows at the GSM authentication and key agreement mechanism. Specifically it shows how the terminals are authenticated with the GSM network.

Unlike the expectations for the 3G network, the GSM authentication mechanism is only one way, therefore the user is not given the assurance that they have established a connection with an authentic serving network.

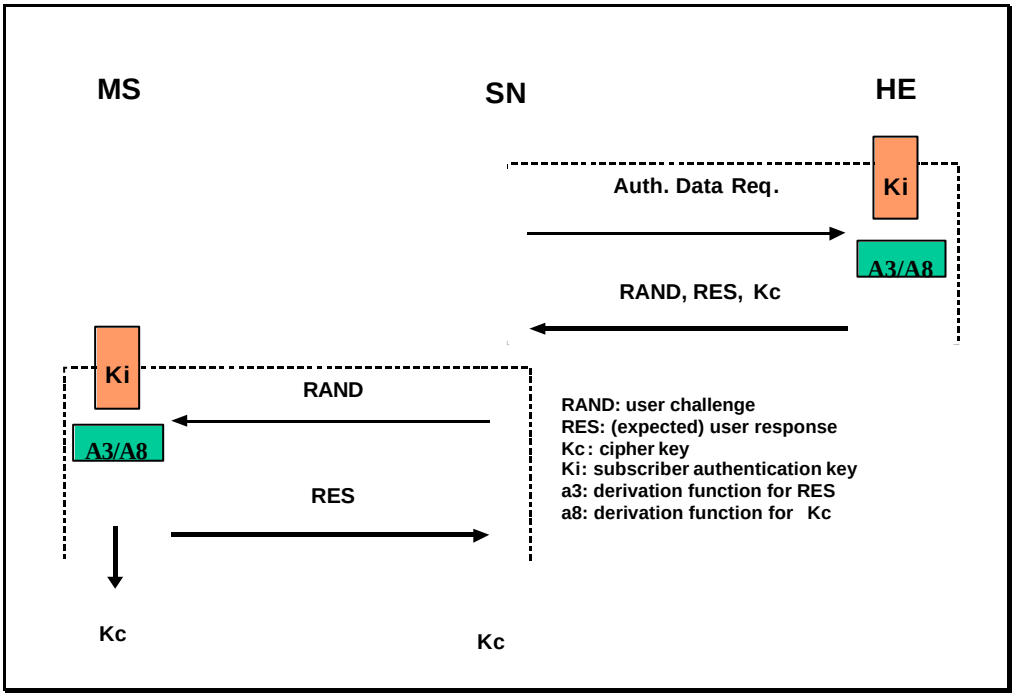

Figure 2 GSM Terminal Authentication using Triplets

The following enhancements to this mechanism were seen as a priority for 3G:

1. Mutual authentication
2. Assurance that authentication information and keys are not being re-used (key freshness)
3. Integrity protection of signalling messages, specifically the secured encryption algorithm negotiation process
4. Use of stronger encryption (a combination of key length and algorithm design)
5. Termination of the encryption further into the core network to encompass microwave links

A new Authentication and Key Agreement (AKA) mechanism, as shown in figure 3, provides the first two of these enhancements. The integrity protection of signalling messages and the algorithm negotiation process are shown in figure 6.

2.4.1 3G Authentication and Key agreement

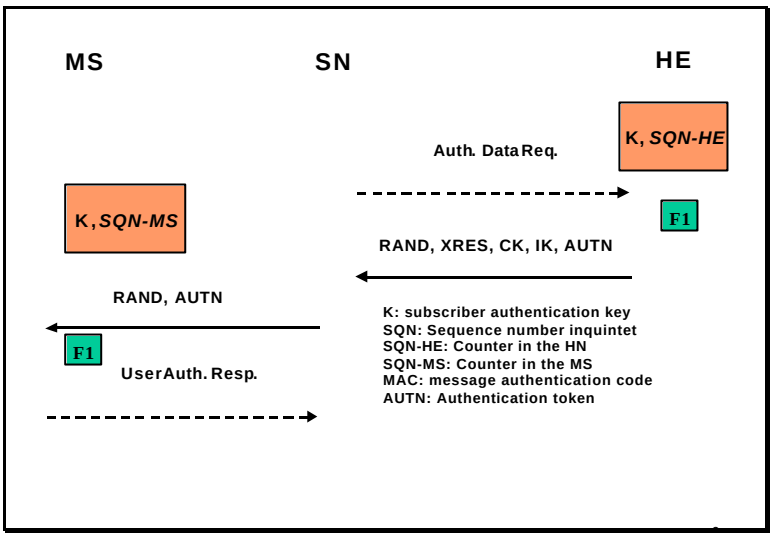

Figure 3 3G Terminal Authentication using Quintets

Additional parameters and cryptographic checks are introduced to provide mutual entity authentication and the establishment of a shared secret cipher key and integrity key between the USIM at the user side and the HLR/AuC at the network side. The mechanism uses symmetric key techniques using a secret subscriber authentication key K that is shared between and available only to the USIM and the HLR/AuC in the user's Home Environment (HE). In addition, the AuC keeps track of a counter SQN_{HE} and the USIM keeps track of a counter SQN_{MS} and stores additional data to support network authentication and to provide the user with assurance of key freshness.

The method is composed of a challenge/response protocol identical to the GSM subscriber authentication and key establishment protocol, combined with a sequence number-based one-pass protocol for network authentication. This protocol was derived from the ISO standard ISO/IEC 9798-4 The HE, which manages both the HLR/AuC and the USIM, has some flexibility in the management of sequence numbers, but some requirements must be met:

a) The mechanism must support secure re-synchronisation of the counter SQN_{HE} in the AuC to the value of the counter SQN_{MS} in the USIM.

b) The mechanism must protect against failures caused by wrap-around of the counter SQN_{MS} in the USIM.

c) The mechanism should not compromise user identity and location confidentiality. Various means of handling this issue are available.

d) The mechanisms for verifying the freshness of sequence numbers in the USIM shall to some extent allow the out-of-order use of sequence numbers. This is to ensure that the authentication failure rate due to synchronisation failures is sufficiently low.

The procedure for distribution of authentication data from the HE to service domain starts with the VLR or SGSN sending a request to the user's HLR/AuC. Upon receipt of that request the HLR/AuC sends an ordered array of *n* quintets (the equivalent of a GSM "triplet") to the VLR or SGSN. To create these quintets the HLR/AuC:

1) Generates a fresh sequence number SQN from a counter SQN_{HE}.

2) Generates an unpredictable challenge RAND.

3) Computes a message authentication code for authentication MAC-A = $f1_K$(SQN || RAND || AMF) where f1 is a message authentication function;

4) Computes an expected response XRES = $f2_K$ (RAND) where f2 is a (possibly truncated) message authentication function;

5) Computes a cipher key CK = $f3_K$ (RAND) where f3 is a key generating function;

6) Computes an integrity key IK = $f4_K$ (RAND) where f4 is a key generating function;

7) Computes an anonymity key AK = $f5_K$ (RAND) where f5 is a key generating function and computes the concealed sequence number SQN ⊕ AK = SQN xor AK. If SQN is to be concealed

8) Assembles the authentication token AUTN = SQN [⊕ AK] || AMF || MAC-A and the quintet Q = (RAND, XRES, CK, IK, AUTN) and updates the counter SQN_{HE}.

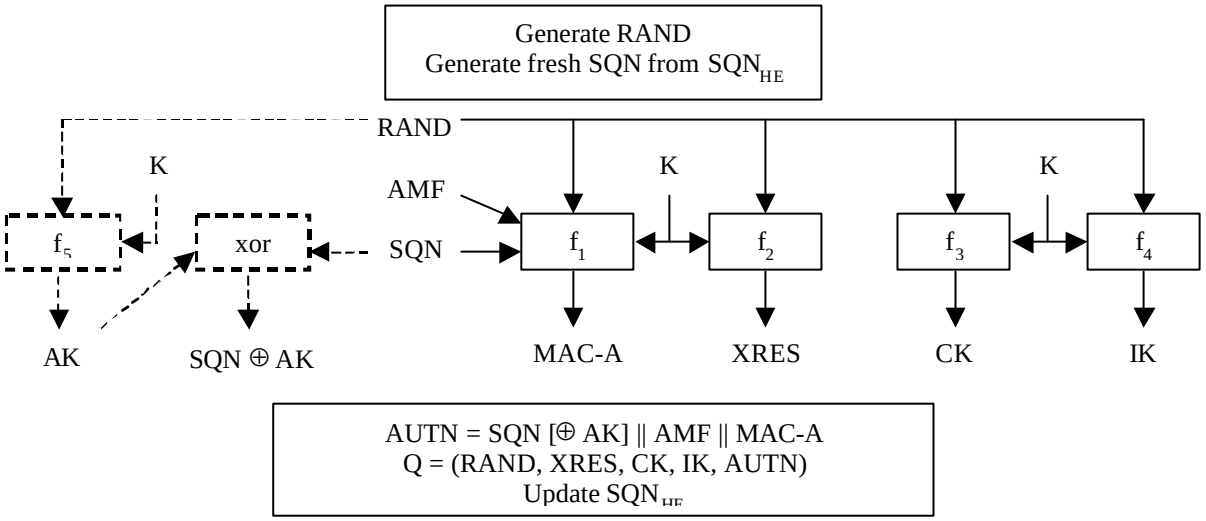

Figure 4: Generation of quintets in the AuC

The Home Environment operator to can use the authentication and key management field AMF included in the authentication token of each quintet to fine tune the performance or bring a new authentication key stored in the USIM into use.

Each quintet consists of the following components:

1. A challenge RAND
2. An expected response XRES
3. A cipher key CK
4. An integrity key IK
5. An authentication token, AUTN = SQN [⊕ AK] || AMF || MAC-A

Each quintet is good for one authentication and key agreement between the VLR or SGSN and the ME/USIM.

1. When the VLR or SGSN initiates the over-the-air authentication and key agreement procedure it selects the next quintet from an array held in the VLR and sends the parameters RAND and AUTN to the user.

2. The USIM checks whether AUTN can be accepted and, if so, produces a response RES that is sent back to the VLR or SGSN. The USIM also computes a session cipher key (CK) and an integrity key (IK).

3. The VLR or SGSN compares the received RES with XRES. If they match the VLR or SGSN considers the authentication and key agreement exchange to be successfully completed and selects the corresponding CK and IK from the quintet.

4. The established keys CK and IK will then be transferred by the USIM and the VLR or SGSN to the entities which perform ciphering and integrity functions, i.e., the ME at the user side and the RNC at the network side.

The processing in the USIM upon receipt of a (RAND, AUTN) is described in detail below. The development of a standard authentication algorithm is being considered in order to provide a minimum level of security for all 3GPP networks: this will still allow some operator-specific variety. The inclusion of variety reduces the risks if the algorithm is compromised, as this will not affect all operators. Operators can also have their own secret algorithm, providing a stronger defence against attacks.

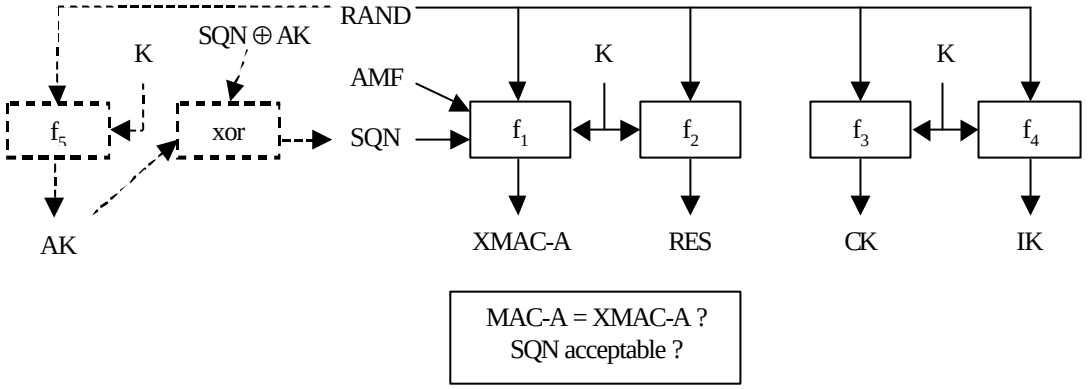

Figure 5: Authentication and key derivation in the USIM

1) If the sequence number is concealed, the USIM computes the anonymity key AK = $f5_K$(RAND) and retrieves from AUTN the unconcealed sequence number SQN = (SQN ⊕ AK) xor AK.

2) The USIM then computes XMAC-A = $f1_K$ (SQN || RAND || AMF) and compares XMAC-A with MAC-A included in AUTN.

3) If they are different, the USIM triggers the ME to send back a *user authentication response* with indication of integrity failure to the VLR or SGSN and abandons the procedure. The next stages are for the case where XMAC-A and MAC-A are equal.

4) Next the USIM verifies that the received sequence number SQN is acceptable. The HE has some flexibility in the management of sequence numbers, but the verification mechanism needs to protect against wrap around and allow to a certain extent the out-of-order use of quintets. There is a detailed description of a mechanism to generate sequence numbers that satisfies all conditions. [Refer to 3G Security Architecture TS33.102] [1]

5) If the sequence number SQN is not acceptable, the USIM computes the re-synchronisation token AUTS and triggers the ME to send back a *user authentication response* back to the VLR or SGSN, with an indication of synchronisation failure, including the re-synchronisation token AUTS, and abandons the procedure. The remaining paragraphs therefore apply for the case where SQN is acceptable. [Refer to 3G Security Architecture TS33.102]

6) The USIM then computes the response RES = $f2_K$(RAND) and triggers the ME to send back a user authentication response back to the VLR or SGSN, with an indication of successful receipt of the signed challenge and including the response RES.

7) Finally the user computes the cipher key CK = $f3_K$ (RAND) and the integrity key IK = $f4_K$ (RAND).

The USIM keeps track of an ordered list of the highest batch number values it has accepted. Using this list mechanism, it is not required that a previously visited SN/VLR deletes the unused authentication vectors when a user de-registers. Retaining the authentication vectors for use when the user returns later may be more efficient as regards signalling. The USIM accepts the sequence number if it is not already on this list or if it is greater than the highest value in the list. In practice, the decision is more complex e.g. we need to prevent forced wrap around of the counter, by ensuring that the USIM will not accept arbitrary jumps in batch numbers. If the sequence number received in an authentication request is accepted, the list is updated. If a sequence number received in a user authentication request is rejected, the list remains unaltered.

2.4.2 Integrity protection of signalling messages

The approach to the negotiation and initialisation of ciphering has been carefully considered with integrity protection applied to the final security mode command that starts the ciphering. The message flows are shown in figure 6. Any intruder attempting to take advantage of users in networks where ciphering is not applied, would have to overcome the mandatory integrity protection to hijack the connection. In networks where ciphering *is* applied the intruder might attempt to spoof a message turning it off, but would be prevented by the integrity protection mechanism.

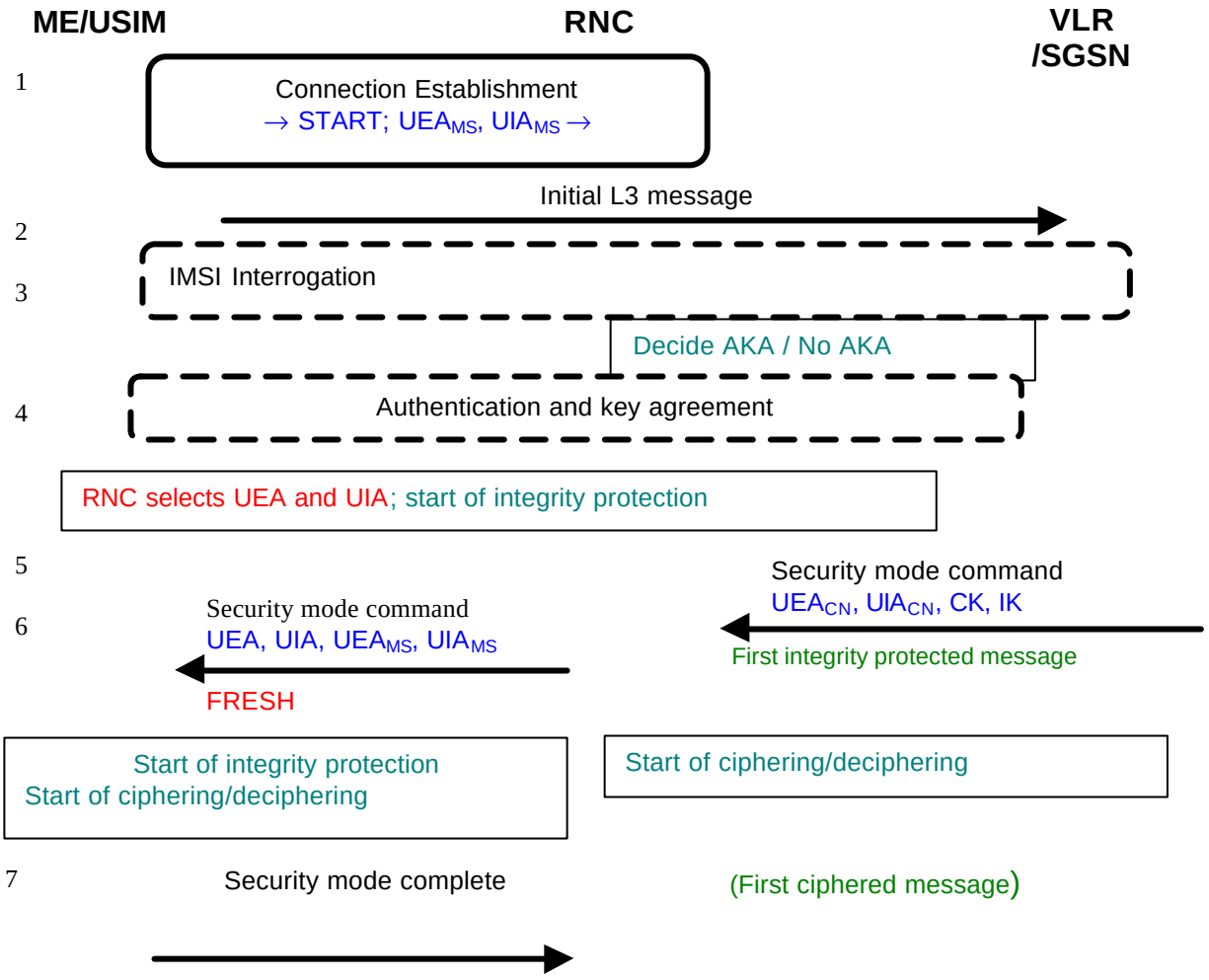

Figure 6 Starting Ciphering & Integrity

2.4.3 Data Integrity on the Air Interface

To protect against false base station attacks, the receiving entity (MS or SN) must be 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). Also, it must be ensured that the data origin of the signalling data received is indeed the one claimed. This is achieved by the inclusion of a data integrity function for signalling data. GSM does not have this functionality. The Message Authentication Code (MAC) function f9 is used to authenticate the data integrity and data origin of signalling data transmitted between the Mobile Equipment (ME) and the Radio Network Controller (RNC). The MAC function f9 is allocated to the ME and the RNC. For 3G Release 99, f9 is based on the Kasumi algorithm.

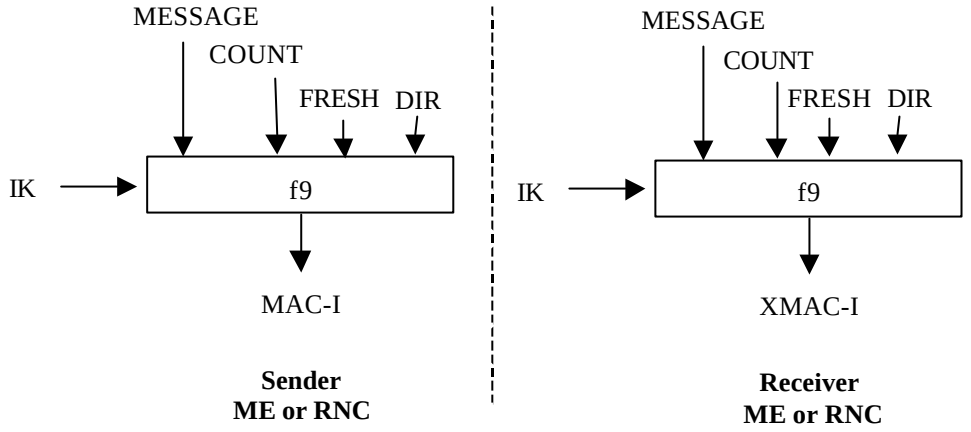

Figure 7 Air Interface Integrity Mechanism

2.4.4 Confidentiality

Protection of User Data is by means of an encryption algorithm, function f8, and used for the protection of user and signalling data sent over the radio access link between RNC and ME. f8 is a symmetric synchronous stream cipher. f8 is used to encrypt plaintext by applying a keystream using a bitwise XOR operation. The plaintext may be recovered by generating the same keystream using the same input parameters and applying it to the ciphertext using a bitwise XOR operation. For 3G Release 99, f8 is based on the Kasumi algorithm.

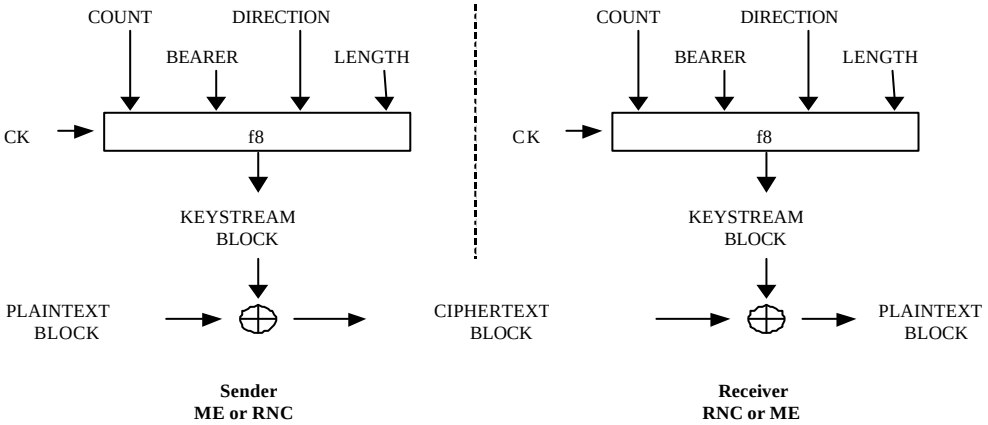

Figure 8 Air Interface Confidentiality Mechanism

3 Conclusion and Outlook

This paper has described the security architecture for 3G that maintains compatibility with GSM as far as possible. Those features of GSM that have proved to be robust and useful have been retained, but enhanced to overcome actual or perceived weaknesses in 2G system. These features are now being designed into Release 99 3G equipment that will be deployed across the world in the next two years. The planning for Release 00 has just commenced and the following security enhancements are planned:

- **Network Domain Security:** This feature ensures that an intruder can inject no malicious operational or maintenance commands into a network domain. The network Security Domain feature provides network elements, in particular network elements belonging to different network operators, with the ability to corroborate each other's identities before exchanging data. A number of options are currently under consideration: a solution for MAP protection at the application layer defined to specify protection of MAP signalling, and a solution for protection of other protocols used for signalling in and between core networks e.g. MAP, CAP, GTP. These mechanisms will require the necessary keys to be established at each involved network element using an automated key management mechanism, standardised to support inter-operation between operators and multi-vendor core networks.

- **IP Multimedia (IM) services:** The IM services will include different applications like voice, video and data. The trust relations and the security services between the end-user, the IM CN subsystem, the PS-domain and the CS-domain will be defined.

- **MExE R00 Enhancements:** MExE is based on the concept of identifying external standards suitable for supporting services from Mobile Equipment (ME), and bringing them into the 3GPP scope by direct reference (i.e. WAP). From the network operator's perspective, it is essential that such developments incorporate security features to preserve the integrity of the network and protect the confidentiality and integrity of third party and end user data and applications.

- **The Open Service Architecture (OSA):** This defines an architecture that enables operator and third party applications to make use of network functionality through an open standardised interface (the OSA Interface). Network/server centric applications can reside outside the core network and make use of service capability features offered through the OSA interface. Applications may also belong to the network operator domain although running outside the core network. From the network operator's perspective, it is essential that such an open interface incorporate security features to preserve the integrity of the network and protect the confidentiality and integrity of third party and end user data and applications.

- **Network Wide Encryption**: The R00 system architecture may create new requirements and/or opportunities for extending confidentiality protection further back into the core network. In addition it may allow for security mechanisms to be applied on an end-to-end basis, providing that the necessary lawful interception requirements are addressed when encryption is applied. This work will take advantage of concepts and hooks for network-wide encryption, which have been considered in R99.

- **Fraud Information Gathering System (FIGS) enhancements:** FIGS provides the means for the Home Network to monitor the activities of its subscribers in a Serving Network. This functionality will need to be extended to cover PS services offered by the PS domain and SIP and H.323 enabled services offered by the IP multimedia (IM) domain.

4 Acknowledgements

The author would like to thank members of 3GPP SA3 for their contributions to the material on which this paper is based, especially to Bart Vinck, Siemens Atea, for figures 2 and 3, and to Guenther Horn, Siemens AG, Peter Howard, Vodafone Limited, and Andrew Myers, BT. Much of the background work on the 3G AKA was done as part of the EU-sponsored collaborative research project USECA [2], the partners are Vodafone Limited, Giesecke & Devrient GmbH, Panasonic PMDC, Siemens Atea, Siemens AG & Katholieke Universiteit Leuven.

5 References

1) Technical Specification 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; 3G Security; Security Architecture (3G TS 33.102 version 3.4.0 Release 1999) (http://www.3gpp.org/)

2) USECA Homepage: (http://www.useca.freeserve.co.uk/)

6 Abbreviations and Symbols

3GPP	Third Generation Partnership Project
AKA	Authentication and key agreement
AMF	Authentication management field
AN	Access Network
AUTN	Authentication Token
AUTS	Re-synchronisation token
AuC	Authentication Centre
AV	Authentication Vector
CN	Core Network
CS	Circuit Switched
FRESH	Random value used for signalling messages replay protection
GGSN	Gateway GPRS Support Node
GSM	Global System for Mobile communications
HE	Home Environment
HLR	Home Location Register
IM	IP Multimedia
IMSI	International Mobile Subscriber Identity
ISO	International Standards Organisation
L3	Layer 3
MAC	Message Authentication Code
MAC-A	The message authentication code included in AUTN, computed using f1
ME	Mobile Equipment
MAP	Mobile Application Part
MExE	Mobile Station Execution Environment
MS	Mobile Station, the combination of ME and USIM (or SIM)
MSC	Mobile Services Switching Centre
OSA	Open Service Architecture
PS	Packet Switched
Q	Quintet, UMTS authentication vector

RAND	Random challenge
RES	(expected) user response to challenge in GSM
RNC	Radio Network Controller
SQN	Sequence number
SQN_{HE}	Sequence number counter maintained in the HLR/AuC
SQN_{MS}	Sequence number counter maintained in the USIM
SGSN	Serving GPRS Support Node
SIM	(GSM) Subscriber Identity Module
SIP	Session Initiation Protocol
SN	Serving Network
T	Triplet, GSM authentication vector
TMSI	Temporary Mobile Subscriber Identity
UEA	UMTS Encryption Algorithm
UIA	UMTS Integrity Algorithm
UICC	UMTS IC Card
USIM	User Services Identity Module
VLR	Visitor Location Register
WAP	Wireless Application Protocol
XRES	Expected Response
XMAC	Expected Message Authentication Code

‖	Concatenation
⊕	Exclusive or
A3	Derivation function for RES user in GSM authentication
A5	Confidentiality algorithm used in GSM
A8	Derivation function for Kc used in GSM
AK	Anonymity Key used in 3G
CK	Cipher Key used in 3G
f1	Message authentication function used to compute MAC
f2	Message authentication function used to compute RES and XRES
f3	Key generating function used to compute CK
f4	Key generating function used to compute IK
f5	Key generating function used to compute AK
f8	3G ciphering function
f9	3G integrity function
IK	Integrity Key used in 3G
K	Long-term secret key shared between the USIM and the AuC in 3G
Kc	Cipher Key used in GSM
Ki	Long-term secret key shared between the SIM and the AuC used in GSM

7 Biography

Colin Blanchard began his career in 1975 as a design engineer with a manufacturer of mobile radio equipment. He joined BT in 1989 and was responsible for a team developing test processes for cellular radio equipment. In 1995, he joined BT laboratories and has acted as the principal security design consultant to a number of BT programmes. From 1996 to 1998 has was Chairman of the SMG10 UMTS Security working group which carried out the initial requirements work and defined the security architecture on which the 3GPP AKA is based. He is currently the BT delegate to the 3GPP SA3 security-working group and to the ETSI TETRA WG6 security-working group. Other areas of interest include the assessment of the security implications of fixed network inter-working with mobile applications and the security impact of Open Services Architecture network API functions.