Stefan Pütz

Security for the Third-Generation Mobile Radio System UMTS

This paper describes the third-generation (3G) security principles specified by the Third-Generation Partnership Project (3GPP) for the third-generation mobile system UMTS as well as the 3G security architecture itself. This is based on five sets of 3G security features and mechanisms. In particular, the differences between second-generation (2G) and 3G security are discussed. The 3G security architecture addresses and corrects all the weaknesses in the 2G systems.

Introduction

Within the last six months a harmonisation and globalisation process for the third-generation mobile radio systems has taken place. Therefore the standardisation process for the universal mobile telecommunications system (UMTS) has moved from the European Telecommunications Standards Institute (ETSI) to the Third-Generation Partnership Project (3GPP). 3GPP is supported by various standards organisations and other related bodies from different continents and countries, for example, Europe (ETSI), Japan (ARIB, TTC), Korea (TTA) and North America (T1), that have agreed to cooperate for the production of a complete set of globally applicable technical specifications for a 3G mobile system based on the evolved global system for mobile communications (GSM) core networks. Third-generation mobile radio systems will involve more players, for example, content and service providers, and more operators, which will result in more roaming. Therefore 3G systems will exist alongside and interact with a lot of different network types. Also 3G systems will promote wireless as the preferred means of communication.

To ensure that UMTS will work securely and reliably under these conditions, a new 3G security architecture has been defined. The scope of 3G security is described by principles. These principles state what is to be provided by 3G security as compared to the security of 2G systems. Security elements within GSM and other 2G systems that have proved to be needed and robust will be adopted. Also 3G security will improve on the security of 2G systems. Therefore 3G security will address and correct real and perceived weaknesses in 2G systems. Last but not least, 3G security will offer new security features and will secure new services offered by 3G mobile radio systems.

Dr. Stefan Pütz:
T-Mobil (DeTeMobil, Deutsche Telekom MobilNet GmbH)
POB 300463, 53184 Bonn, Germany
E-mail: stefan.puetz@t-mobil.de

3G Security Principles

The following three key principles form the basis of 3G security[1]:

- 3G security will build on the security of 2G systems. Security elements within GSM and other 2G systems that have proved to be needed and robust will be adopted for 3G security[4, 5]. Therefore 3G security will retain (and in some cases develop) the following security elements of 2G systems:

  —authentication of subscribers for service access;
  —radio interface encryption†;
  —subscriber identity confidentiality on the radio interface;
  —the SIM as a removable hardware security module that is manageable by network operators and independent of the terminal as regards its security functionality;
  —subscriber identity module (SIM) application toolkit security features providing a secure application layer channel at least between the SIM and a home network server; and
  —home environment (HE) trust in the serving network (SN) for security functionality is minimised.

- 3G security will improve on the security of 2G systems by the way that it will address and correct real and perceived weaknesses in 2G systems as listed below:

  —active attacks based on masquerading a base transceiver station (BTS) are possible;
  —cipher keys and authentication data are transmitted in clear between and within networks;
  —encryption does not extend far enough towards the core network resulting in the cleartext transmission of user and signalling data across microwave links;
  —the provision of protection against channel hijack relies on the use of encryption;

† The strength of the encryption will be greater than that used in 2G systems; the strength is a combination of key length and algorithm design. This is to meet the threat posed by the increased computing power available to those attempting cryptanalysis of the radio interface encryption.

—data integrity is not provided†;
—there is no HE knowledge or control of how an SN uses authentication parameters for HE subscribers roaming in that SN; and
—2G systems do not have the flexibility to upgrade and improve security functionality over time.

- 3G security will offer new security features and will secure new services offered by 3G. More on this issue is discussed in the 'Outlook' section.

As a priority, 3G security will provide the proven 2G security features and correct the weaknesses in 2G systems. Security for new services and service environments will then be developed as required.

3G Security Architecture

The security architecture for the 3G mobile system, specified in Reference 3, consists of five sets of security features and security mechanisms. A security feature is a service capability that meets one or several security requirements. The complete set of security features addresses the security requirements as they are defined in Reference 2. A security mechanism is an element that is used to realise a security feature. All security features and security requirements taken together form the security architecture.

Figure 1 gives an overview of the complete 3G security architecture.

Five security feature groups are defined. Each of these feature groups meets certain threats, and accomplishes certain security objectives. The security feature groups are as follows:

- *network access security (I)*: the set of security features that provide users with secure access to 3G services, and which in particular protect against attacks on the (radio) access link;
- *network domain security (II)*: the set of security features that enable nodes in the provider domain to securely exchange signalling data, and protect against attacks on the wireline network;
- *user domain security (III)*: the set of security features that secure access to mobile stations;
- *application domain security (IV):* the set of security features that enable applications in the user and in the provider domain to securely exchange messages; and
- *visibility and configurability of security (V)*: the set of features that enables the user to inform himself whether a security features is in operation or not and whether the use and provision of services should depend on the security feature.

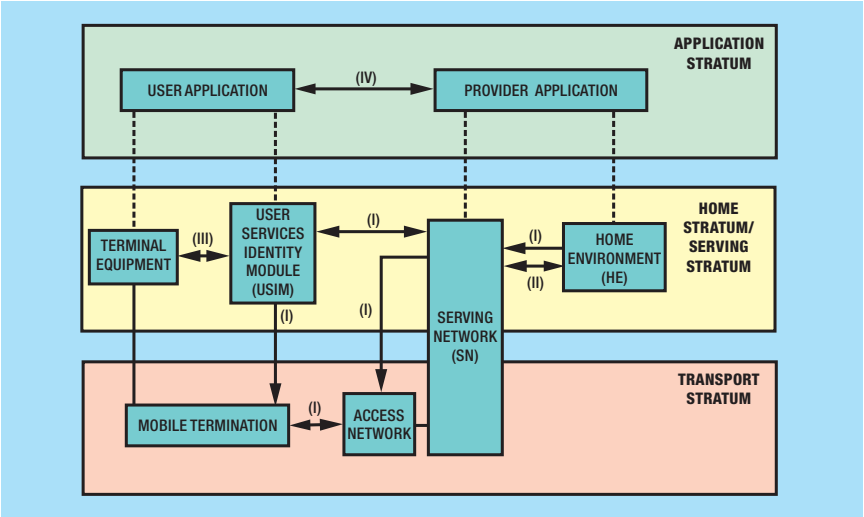

Figure 1—Overview of the security architecture[3]

The following subsections describe the security features belonging to these groups provided by the 3G security architecture.

Network access security

Enhanced user identity confidentiality is the property that the permanent user identity (international mobile user identity (IMUI)) of a user to whom a service is delivered cannot be eavesdropped on the radio access link. Also *user location confidentiality* and *user untraceability* will be provided. Therefore the presence or the arrival of a user in a certain area cannot be determined and 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 by a temporary identity by which he/she is known by the visited serving network, or by an encrypted permanent identity. 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 or encrypted identity. To achieve these security features, 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.

Mutual authentication between user and network will be achieved by the combination of user and network authentication. User authentication is the property that the serving network corroborates the user identity of the user, and network authentication is the property that the user corroborates that he/she is connected to a serving network that is authorised by the user's HE to provide services; this includes the guarantee that this authorisation is recent. To achieve these objectives, it is assumed that authentication should occur at each connection set-up between the user and the network. Two mechanisms have been included: an authentication mechanism using an authentication vector delivered by the user's HE to the serving network, and a local authentication mechanism using the integrity key established between the user and serving network during the previous execution of the authentication and key establishment procedure (see also data integrity).

Confidentiality of user and signalling data is the property that this data cannot be overheard on the radio access interface. This will be achieved by ciphering on the radio interface as well as further back into the network. Cipher key agreement is realised in the course of the execution of the mechanism for authentication and key agreement. Cipher algorithm agreement is

† Data integrity defeats certain false BTS attacks and, in the absence of encryption, provides protection against channel hijack.

realised by means of a mechanism for security mode negotiation between the user and the network. This mechanism also enables the selected ciphering algorithm and the agreed cipher key.

Data integrity and origin authentication of signalling data is the property that the receiving entity is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending entity and that the data origin of the signalling data received is indeed the one claimed. Integrity key agreement is realised in the course of the execution of the mechanism for authentication and key agreement. Integrity algorithm agreement is realised by means of a mechanism for security mode negotiation between the user and the network. This mechanism also enables the selected integrity algorithm and the agreed integrity key.

Network domain security

Network element authentication is the property that a network element corroborates the identity of another network element it wants to communicate with. This feature ensures that no malicious operational or maintenance commands can be injected into a network domain by an intruder. It provides network elements, in particular network elements belonging to different network operators, with the possibility to corroborate each other's identities before exchanging data. This goal may be achieved either by an explicit or implicit entity authentication mechanism, to be performed each time data is exchanged between two network entities.

Confidentiality of exchanged data is the property that data exchanged between two network elements cannot be eavesdropped. If authentication data can be eavesdropped in the network domain, serious fraud problems can arise. Therefore, these features are needed to ensure the confidentiality of sensitive data, for example, authentication or other subscriber data inside the network domain. The features may be realised in the course of an authentication and key agreement mechanism performed by the network elements; the agreed cipher key is then used for securing signalling and user data by means of the cipher algorithm.

Data integrity and data origin authentication of signalling data is the property that the receiving network element is able to verify that signalling data has not been modified in an unauthorised way since it was sent by the sending element and that the data origin of the signalling data received is indeed the one claimed. The feature data integrity of signalling data ensures that operation and maintenance commands or user data exchanged between two network elements cannot be modified by an intruder without being detected, and it ensures that no malicious operational or maintenance commands can be injected into a network domain by an intruder. The features may be realised in the course of an authentication and key agreement mechanism performed by the network entities involved; the agreed integrity key is then used for securing integrity of the exchanged data by means of the integrity algorithm.

User domain security

User-to-USIM authentication provides the property that access to the USIM (UMTS SIM) 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 (for example, a 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 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.

Application security

UMTS is expected to enable operators or third-party providers to create applications which are resident on the USIM or the terminal. Therefore there exists a need to secure messages which are transferred over the 3G network to applications on the USIM or terminal, with the level of security chosen by the network operator or the application provider. *Secure messaging* between the USIM and the network provides a secure application layer channel at least between the USIM and a home network server (similar to the SIM application toolkit in GSM[6, 7]).

Network-wide user traffic confidentiality provides users with the assurance that their traffic is protected against eavesdropping across the entire network, not just on the radio links in the access network, which are particularly vulnerable, but also on the fixed links within the core network.

Security visibility and configurability

Visibility of the operation of security features will be provided in a way that it is transparent to the user if a security feature is enabled or not. This yields to a number of features that inform the user of security-related events, such as:

- indication of access network encryption: the property that the user is informed whether the confidentiality of user data is protected on the radio access link, in particular when non-ciphered calls are set-up;
- indication of network-wide encryption: the property that the user is informed whether the confidentiality of user data is protected along the entire communication path; and
- indication of the level of security: the property that the user is informed on the level of security that is provided by the visited network, in particular when a user is handed over or roams into a network with lower security level (for example 3G to 2G).

Configurability is the property that the user and the user's HE can configure whether the use or the provision of a service should depend on whether a security feature is in operation. A service can only be used if all security features, which are relevant to that service and which are required by the configurations of the user or of the user's HE, are in operation. The following configurability features are suggested:

- enabling/disabling user-USIM authentication: the user and/or user's HE should be able to control the operation of user-USIM authentication; for example, for some events, services or use;
- accepting/rejecting incoming non-ciphered calls: the user and/or user's HE should be able to control whether the user accepts or rejects incoming non-ciphered calls;

Table 1 2G Security Elements Covered by 3G Security Architecture

	Authentication of subscribers	Radio interface encryption	Subscriber identity confidentiality	SIM as a removable hardware security module	SIM application toolkit security features	Minimisation of HE trust in the SN for security functionality
Network access security						
Enhanced user identity confidentiality			x			
Mutual authentication between user and network	x					
Confidentiality of user and signalling data		x				
Data integrity and data origin authentication of signalling data						
User domain security						
Network element authentication						x
Confidentiality of exchanged data						x
Data integrity and data origin authentication of signalling data						x
User domain security						
User-to-USIM authentication				x		
USIM-terminal link				x		
Application security						
Secure messaging					x	
Network-wide user traffic confidentiality						
Security visibility and configurability						
Visibility						
Configurability						
Basic assumption for UMTS				x		

Table 2 2G Security Weaknesses Covered by 3G Security Architecture

	Active attacks based on masquerading BTS	Cipher keys and authentication data are transmitted in clear	Encryption does not extend far enough towards the core network	Protection against channel hijack relies on encryption	Data integrity is not provided	No HE knowledge or control	2G systems do not have enough flexibility
Network access security							
Enhanced user identity confidentiality							
Mutual authentication between user and network	x					x	
Confidentiality of user and signalling data			x				
Data integrity and data origin authentication of signalling data	x			x	x		
User domain security							
Network element authentication							
Confidentiality of exchanged data		x	x				
Data integrity and data origin authentication of signalling data			x			x	
User domain security							
User-to-USIM authentication				x			
USIM-terminal link				x			
Application security							
Secure messaging					x		
Network-wide user traffic confidentiality			x				
Security visibility and configurability							
Visibility							
Configurability							
Basic assumption for UMTS							x

Note for Tables 1 and 2: The first row contains the complete set of security features defined by the 3G security architecture[3]. Columns that are not linked to rows by crosses indicate 3G security features that either do not exist in 2G systems or that are not needed to correct 2G security weaknesses.

- setting up or not setting-up non-ciphered calls: the user and/or user's HE should be able to control whether the user sets up connections when ciphering is not enabled by the network; and
- accepting/rejecting the use of certain ciphering algorithms: the user and/or user's HE should be able to control which ciphering algorithms are acceptable for use.

Mapping Between 2G Security Weaknesses and 3G Security Features

Tables 1 and 2 show that the new 3G security architecture:

- includes all 2G security elements that have been proved to be needed and robust, and
- addresses all real and perceived 2G security weaknesses.

Conclusion

This paper describes in detail the improvement of the new 3G security related to the well known 2G security. It has been shown that all real and perceived 2G security weaknesses are addressed and corrected by the new 3G security features and mechanisms building together the 3G security architecture.

Outlook

Third-generation mobile radio systems will also offer new service capabilities and features that have probably to be secured by additional security features and mechanisms. At the time of writing these cannot be listed here. However, the environment in which these features are likely to be developed can be described. 3G security will secure this environment[1].

The environment in which new services will be developed can be characterised by (but is not limited to) the following aspects. Third-generation mobile radio systems will involve new and different players, for example, content providers or data service providers. These systems will be positioned as the preferred means of communications for users and be preferable to fixed line systems. Probably there will be a variety of prepaid and pay-as-you-go services which may be the rule rather than the exception. A long-term subscription between the user and a network operator may not be the paradigm. Users will have increased control over their service profiles, which they might manage over the Internet and over the capabilities of their terminal. The terminal may support personal authentication of the user, for example, using biometric methods. Also 3G systems will enable new services and functions to be downloaded using extended 2G functions such as MExE (mobile execution environment) and the SIM application toolkit. It is expected that non-voice services will be as important as, or more important than, voice services. Opening the system to the Internet will enable even more services to be used. The terminals will be used as a platform for e-commerce and other applications. Multi-application smartcards, where the USIM is one application among many, can be used with the terminal. The smartcard and terminal will support environments such as Java to allow this. On the other hand there is a higher risk of active attacks† on users that will have to be addressed in the future.

References[*]

1 3G TS 33.120 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Principles and Objectives.
2 3G TS 21.133 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Threats and Requirements.
3 3G TS 33.102 (v3.0.0). Third Generation Partnership Project (3GPP); Technical Specification Group (TSG) SA; 3G Security; Security Architecture.
4 GSM 02.09 (ETS 300 920). Digital Cellular Telecommunications System (Phase 2+); Security Aspects.
5 GSM 03.20 (ETS 300 929). Digital Cellular Telecommunications System (Phase 2+); Security Related Network Functions.
6 GSM 02.48 (TS 101 180). Digital Cellular Telecommunications System (Phase 2+); Security Mechanisms for the SIM Application Toolkit; Stage 1.
7 GSM 03.48 (TS 101 181). Digital Cellular Telecommunications System (Phase 2+); Security Mechanisms for the SIM application toolkit; Stage 2.

† In active attacks, equipment is used to impersonate parts of the network to actively cause lapses in security. In passive attacks, the attacker is outside the system and listens in, hoping security lapses will occur.

* See http://www.etsi.org/ to receive ETSI specifications and http://www.3gpp.org/ to receive 3GPP specifications.

Biography

Stefan Pütz
T-Mobil

Stefan Pütz studied electrical engineering at the University of Siegen, Germany. In 1994, he obtained his diploma. From 1994, he worked on a thesis on 'Non-Deniable Authentication for Future Mobile Radio Systems', carried out at the University of Siegen, Germany. In 1998, he received his Ph.D. in Electrical Engineering. In 1997, he joined Deutsche Telekom MobilNet GmbH, the public network operator in Germany, and now works on the area of systems security in mobile radio networks. Since 1998, he has been a delegate on ETSI SMG10, the group which elaborated the first version of the ETSI description of UMTS security, and in 1999 he was elected as a vice chairman of the 3GPP security group.