Security in 3rd Generation Mobile Networks

Christos Xenakis and Lazaros Merakos

Communication Networks Laboratory
Department of Informatics & Telecommunications
University of Athens, 15784 Athens, Greece.
Tel. +30 210 7275418, +30 210 7275323, Fax. +30 210 7275601
{xenakis,merakos}@di.uoa.gr

ABSTRACT

In the last few years, we have witnessed an explosion in demand for security measures motivated by the proliferation of mobile/wireless networks, the fixed-mobile network convergence, and the emergence of new services, such as e-commerce. 3G-systems play a key role in this network evolution, and thus, all stakeholders are interested in the security level supported in the new emerging mobile environment. This paper elaborates on the security framework in 3G mobile networks. The security requirements imposed by the different types of traffic, and by the different players involved (mobile users, serving network and service providers) are investigated. The security architecture, which comprises all the security mechanisms that are projected for the Universal Mobile Telecommunication System (UMTS) network, is analyzed. The employment of traditional security technologies, originally designed for fixed networking, such as firewalls, and static Virtual Private Network (VPN), in order to safeguard the UMTS core network from external attacks, as well as to protect user data when conveyed over the network are examined. Critical points in the 3G-security architecture that may cause network and service vulnerability are identified and discussed. Furthermore, proposals for the enhancement of the 3G-security architecture, and the provision of advanced security services to end-user data traffic within and outside the UMTS core network are discussed. The proposed enhancements can be easily integrated in the existing network infrastructure, and operate transparently to the UMTS network functionality.

1. Introduction

Mobile/wireless Internet is becoming available with the advent of third generation (3G) mobile communication systems. Along with the variety of new perspectives, mobile Internet also raises new concerns on security issues. Wireless access is inherently less secure, and mobility implies higher security risks compared to those encountered in fixed networks. The advanced wireless and wired network infrastructure, which supports higher access rates, and the complex network topologies, which enable “anywhere-anytime” connectivity, may increase the number and the ferocity of potential attacks. Furthermore, the potential intruders are able to launch malicious attacks from mobile devices with enhanced processing capabilities, which are difficult to trace.

The introduction of IP-based transport technology to the core of 3G mobile networks brings along new vulnerabilities and potential threats. Presently, mobile network operators do not deploy their own private networks, but they rather rely on the existing Internet infrastructure for the establishment of intra-network and inter-network communications. The major security concern in this approach is that the “shared” Internet links are exposed to a large population of aggressors, who have direct access to them.

To counteract against these vulnerable points, 3G-systems have incorporated a specific security architecture. This architecture is built on the security principles of 2G-systems with improvements and enhancements in certain points in order to provide advanced security services. Its main objective is to ensure that all information generated by or relating to a user, as well as the resources and services provided by the serving network (SN) and the home environment (HE), are adequately protected against misuse or misappropriation.

This paper examines and analyses the security framework in 3G mobile networks. The security requirements imposed by the different types of traffic, and by the different players involved (mobile users, serving network and service providers) are investigated. The security architecture, which comprises all the security mechanisms that are projected for the Universal Mobile Telecommunication System (UMTS) network [2], is presented and elaborated. The employment of traditional security technologies, originally designed for fixed networking, such as firewalls, and static VPN, in order to safeguard the UMTS core network from external attacks, as well as to protect user data when are conveyed over the network are examined. Critical points in the 3G-security architecture that may cause network and service vulnerability are identified. Next generation mobile subscribers require generic, client initiated security mechanisms, which can provide advanced security services to user data traffic according to the particular end-users needs, and will be available anywhere – anytime. Evolution perspectives that aim to enhance the 3G-security architecture, as well as to provide advanced security services to user data traffic within and outside the UMTS core network are analyzed. These proposed improvements can be easily integrated in the existing network infrastructure and operate transparent to the UMTS network functionality.

The rest of this paper is organized as follows. Section 2 outlines the UMTS network architecture, and the security requirements imposed by the involved players. Section 3 elaborates on the network access security features. Section 4 examines the network domain security, as well as the employment of firewalls and static VPNs technology in the 3G-network architecture. Section 5 presents the user domain security, the application domain security, the visibility of security operation and configurability, and the network-wide confidentiality option. Section 6 analyses potential weaknesses concerning the 3G-security architecture, and presents evolution perspectives that aim to enhance the supported security framework. Finally, section 7 contains the conclusions.

2. UMTS network

2.1 Network architecture

UMTS comprises a realization of 3G-mobile systems, which is compatible with the evolved Global System for Mobile communication / General Packet Radio Services (GSM/GPRS) [1] network. UMTS has been standardized in several releases, starting from Release 1999 (R99) and moving forward to Release 4 (Rel-4), Release 5 (Rel-5) etc. Its main objective is to provide a wide range of real-time multimedia applications with differentiated levels of quality of service and advanced service features to mobile users. The fundamental difference between GSM/GPRS and UMTS R99 is that the latter supports higher access rates (up to 2Mbps) [2]. This is achieved through a new Wideband Code Division Multiple Access (WCDMA) radio interface for the land based communication system, named UMTS Terrestrial Radio Access Network (UTRAN) [3]. Fig. 1 depicts the UMTS generic network architecture.

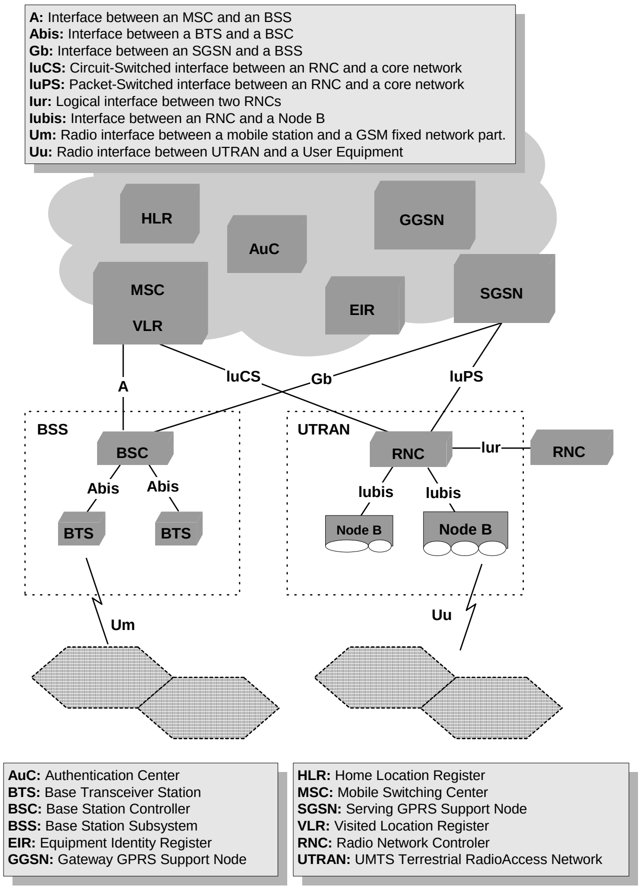
A: Interface between an MSC and an BSS
Abis: Interface between a BTS and a BSC
Gb: Interface between an SGSN and a BSS
IuCS: Circuit-Switched interface between an RNC and a core network
IuPS: Packet-Switched interface between an RNC and a core network
Iur: Logical interface between two RNCs
Iubis: Interface between an RNC and a Node B
Um: Radio interface between a mobile station and a GSM fixed network part.
Uu: Radio interface between UTRAN and a User Equipment
HLR
AuC
GGSN
MSC
VLR
EIR
SGSN
A
IuCS
Gb
IuPS
BSS
BSC
Abis
Abis
BTS
BTS
UTRAN
RNC
Iur
RNC
Iubis
Iubis
Node B
Node B
Um
Uu
AuC: Authentication Center
BTS: Base Transceiver Station
BSC: Base Station Controller
BSS: Base Station Subsystem
EIR: Equipment Identity Register
GGSN: Gateway GPRS Support Node
HLR: Home Location Register
MSC: Mobile Switching Center
SGSN: Serving GPRS Support Node
VLR: Visited Location Register
RNC: Radio Network Controler
UTRAN: UMTS Terrestrial RadioAccess Network

Fig. 1: UMTS network architecture

While UMTS R99 is a logical evolution from the 2G-system architecture, UMTS Rel-4 and Rel-5 are revolutionary, introducing new concepts and advanced features [4, 5]. A major point of differentiation is the shift towards an all-IP network architecture that eventually will replace the circuit-switched (CS) transport technology, which is partially used in UMTS R99, by packet-switched (PS). Another difference is the incorporation of an Open Service Architecture (OSA), which allows network operators to provide third party access to their UMTS service architecture. Therefore, the evolution of UMTS network architecture signifies not only a shift towards a common IP-based platform, which guarantees interworking with existing and forthcoming networks, but also a shift towards an open and easily accessible network.

2.2 Security architecture and requirements

Security protection in 3G-networks requires the consideration of several aspects and issues, such as the wireless access, the end-user mobility, the particular security threats, the type of information to be protected, and the complexity of the network architecture. The radio transmission is by nature more susceptible to eavesdropping and fraud in use than wireline transmission. The user mobility and the universal network access certainly provoke security treats. The different types of data, such as user data, charging and billing data, customer information data, and network management data, which are conveyed or are resident in mobile networks, require different type and level of protection. Furthermore, the complex network topologies and the heterogeneity of the involved technologies increase the dependability challenge.

Fig. 2 gives an overview of the complete 3G-security architecture, illustrating five major security classes: (I) network access security, (II) network domain security, (III) user domain security, (IV) application domain security and (V) visibility and configurability of security [13]:

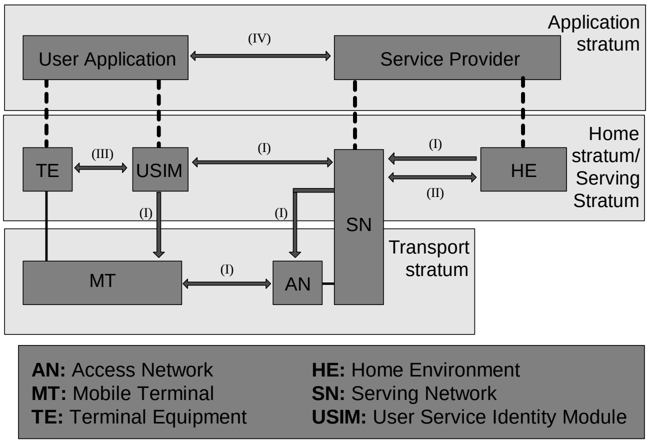

Fig. 2 3G-security architecture

Although mobile networks differ in nature from fixed terrestrial networks, their security measures should also support the principles defined for traditional IP networking, such as confidentiality, integrity, authentication, availability, authorization and accounting [6]. These measures counteract against a number of potential attacks like masquerading, unauthorized use of resources, unauthorized disclosure of information, unauthorized alteration of information, repudiation of actions and denial of service [7, 8].

Specifically, a mobile user connected to a 3G-network should be able to verify that the SN is authorized to offer services on behalf of the user’s HE at the start of, and during, the service delivery. All data exchange, occurring between the mobile user and the SN or the service provider (SP), must be protected against unauthorized modification. Moreover, the mobile user should be capable of checking whether data traffic and call-related information is confidentially protected. The end-user has also to be assured that no personal information, such as user identity or user location, is revealed to other individuals.

From the SN point of view, any potential intruder should be prevented from obtaining unauthorized access to services by masquerading as an authorized user. It must be possible for the HE to immediately terminate all services provided to a certain user or group of users, in case they break the service offering rules. The SN has to be able to authenticate the origin of user traffic, signaling, and control data, especially over the vulnerable radio interface. Moreover, the network has to protect the confidentiality as well as the unauthorized modification of user data, signaling and control data, which either reside in the network, or travel through it.

Finally, the SP has to authenticate the users at the start of and during the service delivery, in order to prevent intruders from obtaining unauthorized access. Furthermore, the SP must be able to detect and prevent the fraudulent use of services (e.g., unauthorized access to data while being downloaded to an authorized user).

In the following, the security features included in the 3G-security architecture aiming at satisfying the security requirements of the involved communicating parties (end-users, SN and SP) are presented and analyzed.

3. Network access security

Network access security is a key component in the 3G-security architecture. This class deals with the set of security mechanisms that provide users with secure access to 3G services, as well as protect against attacks on the radio interface. Such mechanisms include: i) user identity confidentiality, ii) authentication and key agreement iii) data confidentiality and iv) integrity protection of signaling messages. Network access security takes place independently in each service domain.

3.1 User identity confidentiality

User identity confidentiality allows the identification of a user on the radio access link by means of a Temporary Mobile Subscriber Identity (TMSI). This implies that confidentiality of the user identity is protected almost always against passive eavesdroppers. Initial registration is an exceptional case where a temporary identity cannot be used, since the network does not yet know the permanent identity of the user.

The allocated temporary identity is transferred to the user once the encryption is turned on. A TMSI in the CS domain or P-TMSI in PS domain has a local significance only in the location area or the routing area, in which the user is registered. The association between the permanent and temporary user identities is stored in the Visited Location Register or the Serving GPRS Support Node (VLR/SGSN). If the mobile user arrives into a new area, then, the association between the permanent and the temporary identity can be fetched from the old location or routing area. If the address of the old area is not known or the connection cannot be established, then, the permanent identity must be requested from the mobile user.

To avoid user traceability, which may lead to the compromise of user identity confidentiality as well as to user location tracking, the user should not be identified for a long period by means of the same temporary identity. Additionally, any signaling or user data that might reveal the user's identity are ciphered on the radio access link.

3.2 Authentication and key agreement

Authentication and key agreement mechanism achieves mutual authentication between the mobile user and the SN showing knowledge of a secret key K, as well derives ciphering and integrity keys. The authentication method is composed of a challenge/response protocol (see Fig. 3), and was chosen in such a way as to achieve maximum compatibility with the GSM/GPRS security architecture facilitating the migration from GSM/GPRS to UMTS. Furthermore, the User Service Identity Module (USIM) [14] and the HE keep track of counters SQN_{MS} and SQN_{HE}, respectively, to support the network authentication. The sequence number SQN_{HE} is an individual counter for each user, while the SQN_{MS} denotes the highest sequence number that the USIM has accepted.

Upon receipt of a request from the VLR/SGSN, the HE authentication center (HE/AuC) forwards an ordered array of authentication vectors (AV) to the VLR/SGSN. Each AV, which is used in the authentication and key agreement procedure between the VLR/SGSN and the USIM, consists of a random number RAND, an expected response XRES, a cipher key CK, an integrity key IK, and an authentication token AUTN.

Fig. 4 shows an AV generation by the HE/AuC. The HE/AuC starts with generating a fresh sequence number SQN, which proves to the user that the generated AV has not been used before, and an unpredictable challenge RAND. Then, using the secret key K it computes:

- The Message Authentication Code *MAC = f1k (SQN ||*[1] *RAND || AMF),* where f1 is a message authentication function, and the Authentication and key Management Field (AMF) is

[1] || String concatenation

used to fine tune the performance or bring a mew authentication key stored in the USIM into use.

- The expected response $XRES = f2k\ (RAND)$ where f2 is a (possibly truncated) message authentication function.
- The Cipher Key $CK = f3k\ (RAND)$,
- the Integrity Key $IK = f4k\ (RAND)$,
- and the Anonymity Key $AK = f5k\ (RAND)$ where f3, f4 and f5 are key generating functions.
- Finally, the HE/AuC assembles the authentication token $AUTN = SQN \oplus$[2] $AK \,||\, AMF \,||\, MAC$

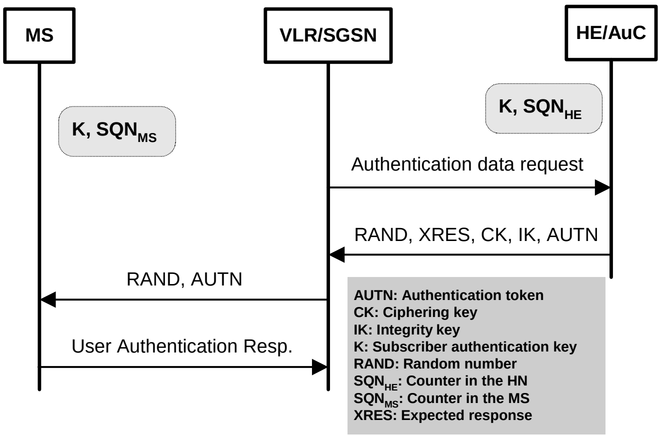
MS
VLR/SGSN
HE/AuC
K, SQN_HE
K, SQN_MS
Authentication data request
RAND, XRES, CK, IK, AUTN
RAND, AUTN
User Authentication Resp.
AUTN: Authentication token
CK: Ciphering key
IK: Integrity key
K: Subscriber authentication key
RAND: Random number
SQN_HE: Counter in the HN
SQN_MS: Counter in the MS
XRES: Expected response

Fig. 3: 3G authentication and key agreement

It has to be noted that the authentication and key generation functions f1, f2, f3, f4 and f5, and the consequent AV computation follow the one-way property. This means that if the output is known there exist no efficient algorithm to deduce any input that would produce the output. Although the f1-f5 functions are based on the same basic algorithm, they differ from each other in a fundamental way in order to be impossible to deduce any information about the output of one function from the output of the others. Since they are used in the AuC and in the USIM, which are controlled by the home operator, the selection of the algorithms (f1-f5) is in principal operator specific. However, an example algorithm set has been proposed called MILENAGE [23].

When the VLR/SGSN initiates an authentication and key agreement procedure, it selects the next AV from the ordered array, and forwards the parameters RAND and AUTN to the user. The USIM using also the secret key K computes the AK,

$AK = f5k\ (RAND)$,

and retrieves the SQN,

$SQN = (SQN \oplus AK) \oplus AK$.

Then, it computes $XMAC = f1k\ (SQN \,||\, RAND \,||\, AMF)$, and checks whether the received AUTN and the retrieved SQN values were indeed generated in AuC [13].

If so, the USIM computes the $RES = f2k\ (RAND)$, and triggers the mobile station (MS) to send back a user authentication response. Afterwards, the USIM computes the CK,

[2] ⊕ Exclusive or

CK = f3k (RAND),
and the IK,
IK = f4k (RAND).

The VLR/SGSN compares the received RES with the XRES field of the AV. If they match, it considers that the authentication and key agreement exchange has been successfully completed. Finally, the USIM and the VLR/SGSN transfer the established encryption and integrity protection keys (CK and IK) to the mobile equipment and the Radio Network Controller (RNC) that perform ciphering and integrity functions.

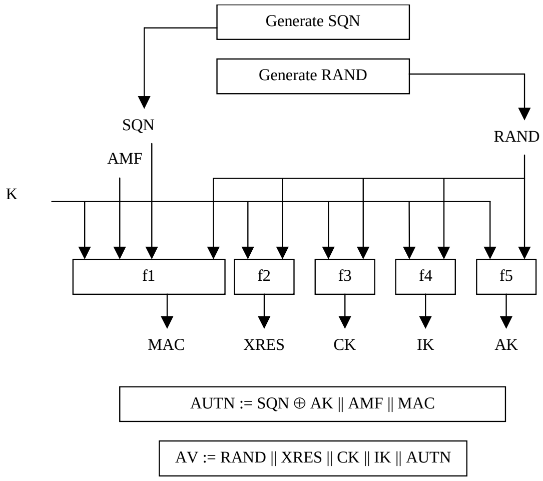
Generate SQN
Generate RAND
SQN
RAND
AMF
K
f1
f2
f3
f4
f5
MAC
XRES
CK
IK
AK
AUTN := SQN ⊕ AK || AMF || MAC
AV := RAND || XRES || CK || IK || AUTN

Fig. 4: Generation of authentication vectors

3.3 Data confidentiality

Once the user and the network have authenticated each other, they may begin secure communication. As described above, a cipher key is shared between the core network and the terminal after a successful authentication event. User and signaling data sent over the radio interface are subject to ciphering using the function f8. The encryption/decryption process takes place in the MS and the RNC on the network side. The f8 is a symmetric synchronous stream cipher algorithm that is used to encrypt frames of variable length. The main input to the f8 is a 128-bit secret cipher key CK. Additional inputs, which are used to ensure that two frames are encrypted using different keystreams, are a 32-bit value COUNT, a 5-bit value BEARER and an 1-bit value DIRECTION (fig.5). The output is a sequence of bits (the 'keystream') of the same length as the frame. The frame is encrypted by XORing the data with the keystream. For UMTS R99, f8 is based on the Kasumi algorithm [15].

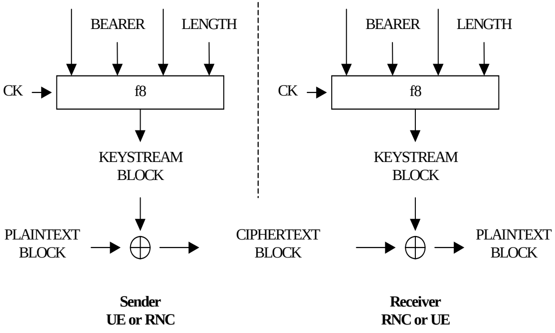

Fig. 5: Ciphering over the radio access link.

3.4 Integrity protection of signaling messages

The radio interface in 3G-mobile systems has also been designed to support integrity protection on the signaling channels. This enables the receiving entity to be able to verify that the signaling data has not been modified in an unauthorized way since it was sent. Furthermore, it ensures that the origin of the received signaling data is indeed the one claimed. The integrity protection mechanism is not applied for the user plane due to performance reasons.

The function f9 is used to authenticate the integrity and the origin of signaling data between the MS and the RNC in UMTS. It computes a 32-bit Message Authentication Code (MAC) (Fig. 6), which is appended to the frame, and is checked by the receiver. The main inputs to the algorithm are an 128-bit secret integrity key IK, and the variable-length frame content MESSAGE. Additional inputs, which are used to ensure that MACs for two frames with identical content are different, are a 32-bit value COUNT, a 32-bit value FRESH, and an 1-bit value DIRECTION. In the UMTS R99, the f9 is based on the Kasumi algorithm [15].

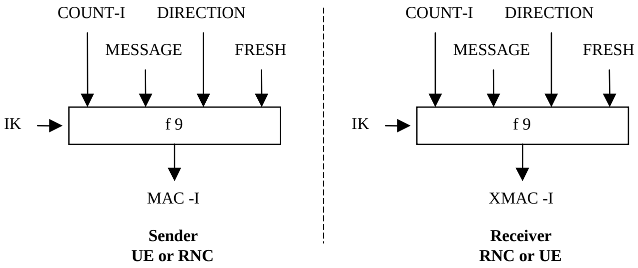

Fig. 6: Derivation of MAC on a signaling message.

4 Network domain security

Network domain security (NDS) features ensure that signaling exchanges within the UMTS core, as well as in the whole wireline network are protected. Various protocols and interfaces are used for the control plane signaling inside, and between core networks, such as the Mobile Application Part (MAP) [16] and the GPRS Tunneling Protocol (GTP) [1] protocols, and the Iu (IuPS, IuCS) and Iur [2] interfaces. These will be protected by standard procedures based on the existing cryptographic techniques. Specifically, the IP–based protocols shall be protected at network level by means of IPsec [6], while the realization of protection for the SS7-based protocols and the lu and Iur interfaces shall be accomplished at the application layer. In the following, the NDS context for IP-based [17] and SS7-based [18] protocols is presented. Furthermore, the employment of traditional security technologies, originally designed for fixed networking, such as firewalls, and static VPN, in order to safeguard the UMTS core network from external attacks, as well as to protect user data when are conveyed over the public internet are examined.

4.1 IP-based protocol

The UMTS network domain control plane is sectioned into security domains, which typically coincide with the operator borders. Security gateways (SEGs) are entities at the borders of the IP security domains used for securing native IP-based protocols. It is noted that NDS does not extend to the user plane, which means that packet flows over the Gi [2] interface will not be protected by the SEGs. The key management functionality is logically separate from the SEG. Key administration centers (KACs) negotiate the IPsec security associations (SAs) by using the Internet Key Exchange (IKE) protocol [19] in a client mode, on behalf of the network entities (NEs) and the SEGs. The KACs also distribute SAs parameters to the NEs or the SEGs through standard interfaces. In Fig. 7 the UMTS NDS architecture for IP-based protocols is depicted.

To secure the IP traffic between two NEs, either a hop-by-hop or an end-to-end scheme may be applied. The first requires the originating NE to establish an IPsec tunnel to the appropriate SEG in the same security domain and forward the data to it. The SEG terminates this tunnel and sends the data through another IPsec tunnel to the receiving network. The second tunnel is terminated by the SEG in the receiving domain, which in turn uses IPsec to pass the data to its final destination (path (a) in Fig. 7). The end-to-end scheme implies that an IPsec SA is established between the two NEs (path (b) in Fig. 7). This scheme can also be applied in case that the two parties belong to the same security domain.

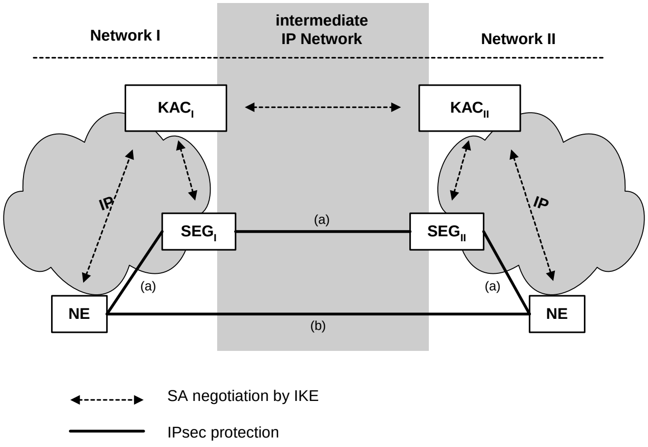

Fig. 7: NDS architecture for IP-based protocols

Node authentication can be accomplished using either pre-shared symmetric keys, or public keys [19]. Using pre-shared symmetric keys means that the KACs or the NEs do not have to perform public key operations, as well as there is no need for establishing a public key infrastructure. The IPsec can be configured either in transport mode or in tunnel mode [6]. Whenever at least one endpoint is a gateway, then, the tunnel mode suits better. Finally, the IPsec protocol shall always be Encapsulation Security Payload (ESP) [6], given that it can provide confidentiality and integrity protection as well.

4.2 SS7-based protocols

NDS for SS7-based protocols is mainly found at the application layer. Specifically, in case that the transport relies on SS7, or on a combination of SS7 and IP, then, security shall be provided at the application layer. On the other hand, whenever the transport is based on IP only, security may be provided at the network layer exclusively, or in addition to the application layer security, by using IPsec. For signaling protection at the application layer the necessary SAs will be network-wide and they are negotiated by KAC, similarly to the IP-based architecture (Fig.8). End-to-end protected signaling will be indistinguishable to unprotected signaling traffic to all parties, except for the sending and receiving sides.

It is worth noticing that in Rel-4 the only protocol that is to be protected is the MAP. The complete set of enhancements and extensions that facilitate the MAP security is termed MAPsec [18]. The MAPsec covers the security management procedures, as well as the security of the transport protocol including data integrity, data origin authentication, anti-reply protection and confidentiality. Finally, for IKE adaptation a specific Domain of Interpretation is required.

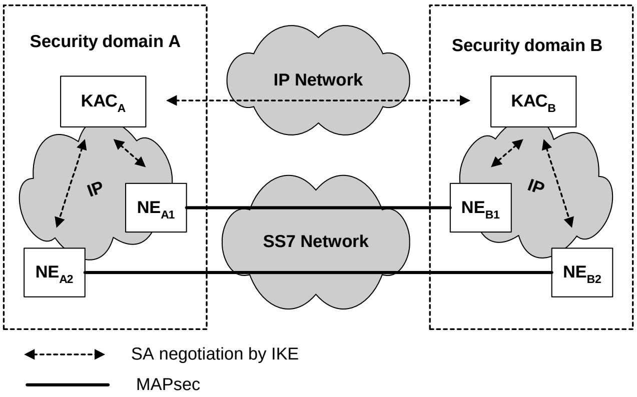

Fig. 8: NDS architecture for SS7 and mixed SS7/IP-based protocols

4.3 Traditional network security features

Besides the security features that are included in the 3G-security architecture, the mobile network operators can apply traditional security technologies used in terrestrial networking to safeguard the UMTS core network, as well as the inter-network communications. User data in the UMTS backbone network are conveyed in clear-text exposing them to various external threats. Moreover, inter-network communication is based on the public Internet, which enables IP spoofing to any malicious third party who gets access to the network. Thus, to counteract against these vulnerable points the mobile operators can use two complementary technologies: firewalls, and VPNs [21].

Firewalls can be characterized as a technology providing a set of mechanisms to enforce a security policy on data from and to a corporate network. They are established at the borders of core network allowing traffic originated from specific foreign IP addresses. Thus, firewalls protect the UMTS backbone from unauthorized penetration. Furthermore, application firewalls prevent direct access through the use of proxies for services, which analyze application commands, perform authentication, and keeps logs.

Since firewalls do not provide privacy and confidentiality, VPNs have to complement them to protect data in transit. VPN establishes a secure tunnel between two points, encapsulates and encrypts data, and authenticates and authorizes user access of the corporate resources on the network. Thus, they extend dedicated connections between remote branches, or remote access to mobile users, over a shared infrastructure. Implementing a VPN makes security issues such as confidentiality, integrity, and authentication, paramount. There is a two-fold benefit that arises from VPN deployment: the low cost and security.

A number of different mechanisms that provide VPN over IP networks exist. Application-layer security builds security features into individual applications, and they operate independently of any network security measures. Many applications have special security requirements that simply cannot be met by network security services. Security at this level is by far the easiest to deploy, as long as all users are running a homogeneous application on a standard platform. While these methods are effective for solving specific security problems, such solutions are by their nature limited to their specific niches. For instance, Transport Layer Security (TLS) [21] works fine for simple client-service cases, but, in case that the service contains a considerable number of cross-references to other servers, for each one a separate key-exchange operations is needed overloading unnecessarily the client.

Link-layer techniques are often applied within wireless link. Their limitation pertains to the fact that link-layer end-to-end security cannot be consistently realized in the context of a network utilizing various types of radio and wired transmission segments. On the other hand, the IPsec [6] standard aims at securing the network layer, and guarantees security for any application that uses it. It facilitates transparent encryption and integrity protection on both IPv4 and IPv6, and authentication of the communicating peers. It is especially useful for implementing VPNs, and for remote access to private networks.

The border gateway is an element that resides at the border of the UMTS core network and provides the appropriate level of security policy (e.g., firewall), as well as maintaining static pre-configured security tunnels (e.g., IPsec tunnels) granting VPN services to specific peers. It serves as a gateway between the PS domain and an external IP network that is used to provide connectivity with other PS domains located in other core networks. The border gateway is required only to support PS-type of services.

5. User and application domain security features

5.1 User domain security

User domain security [13] ensures secure access to the MS. It is based on a physical device called UMTS Integrated Circuit Card (UICC), which can be easily inserted and removed from terminal equipment, containing security applications such as the USIM [14]. The USIM represents and identifies a user and his association to an HE. It is responsible for performing subscriber and network authentication, as well as key agreement, when 3G services are accessed. It may also contain a copy of the user’s profile.

The USIM access is restricted to an authorized user, or to a number of authorized users. To accomplish this feature, the user and the USIM must share a secret (e.g., a PIN). The user gets access to the USIM only if he proves knowledge of the secret. Furthermore, access to a terminal or to other user equipment can be restricted to an authorized USIM. To this end, the USIM and the terminal must also share a secret. If a USIM fails to prove its knowledge of the secret, then, access to the terminal is denied.

5.2 Application domain security

On the other hand, application domain security [13] deals with secure messaging between the MS and the SN or the SP over the network with the level of security chosen by the network operator or the application provider. A remote application should authenticate a user before allowing him to utilize the application services, and it could also provide for application-level data confidentiality. Application-level security mechanisms are needed because the lower layers’ functionality may not guarantee end-to-end security provision. Lack of end-to-end security could be envisioned when, for instance, the remote party is accessible through the Internet.

USIM Application Toolkit [20] provides the capability for operators or third party providers to create applications that are resident on the USIM. To assure secure transactions between the MS and the SN or the SP, a number of basic security mechanisms such as entity authentication, message authentication, replay detection, sequence integrity, confidentiality assurance, and proof of receipt, have been specified and integrated in the USIM Application Toolkit.

Wireless Application Protocol (WAP) is a suite of standards for delivery and presentation of Internet services on wireless terminals, taking into account the limited bandwidth of mobile networks, as well as the limited processing capabilities of mobile devices. To connect the wireless domain to the Internet, a WAP gateway is needed to translate the protocols used in WAP segment to the protocols used in the public Internet. The WAP architecture has been standardized in two releases (ver. 1.2.1 and ver. 2.0) [24].

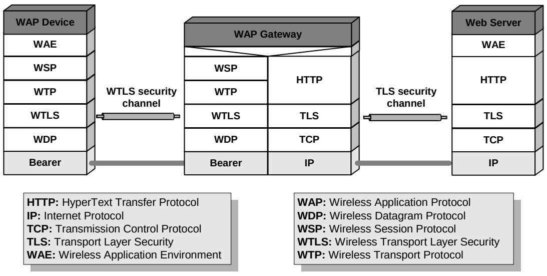
WAP Device
WAE
WSP
WTP
WTLS
WDP
Bearer
WTLS security channel
WAP Gateway
WSP
WTP
WTLS
WDP
Bearer
HTTP
TLS
TCP
IP
TLS security channel
Web Server
WAE
HTTP
TLS
TCP
IP
HTTP: HyperText Transfer Protocol
IP: Internet Protocol
TCP: Transmission Control Protocol
TLS: Transport Layer Security
WAE: Wireless Application Environment
WAP: Wireless Application Protocol
WDP: Wireless Datagram Protocol
WSP: Wireless Session Protocol
WTLS: Wireless Transport Layer Security
WTP: Wireless Transport Protocol

Fig. 9.a: WAP 1.2.1 architecture

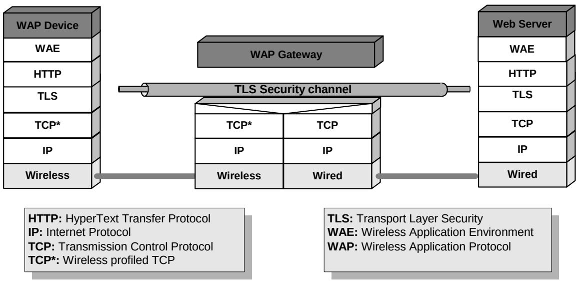
WAP Device
WAE
HTTP
TLS
TCP*
IP
Wireless
WAP Gateway
TLS Security channel
TCP*
IP
Wireless
TCP
IP
Wired
Web Server
WAE
HTTP
TLS
TCP
IP
Wired
HTTP: HyperText Transfer Protocol
IP: Internet Protocol
TCP: Transmission Control Protocol
TCP*: Wireless profiled TCP
TLS: Transport Layer Security
WAE: Wireless Application Environment
WAP: Wireless Application Protocol

Fig. 9.b: WAP 2.0 architecture

To secure data transmission in the WAP architecture (ver. 1.2.1), the Wireless Transport Layer Security (WTLS) protocol [24], which is based upon the TLS protocol, is employed. WTLS has been optimized for use over narrow-band communication channels providing also datagram support. It ensures data integrity, privacy, authentication, and denial-of-service protection. For Web applications that employ standard Internet security techniques with TLS, the WAP gateway automatically and transparently manages wireless security, and conveys protected data between the WTLS and TLS security channels (see Fig. 9.a). Thus, this scheme does not support end-to-end security.

WAP 2.0 proceeds to the re-design of the WAP architecture by introducing the existing Internet protocol stack, including the Transmission Control Protocol (TCP), into the WAP environment. The new architecture allows a range of different gateways, which enables conversion between the two protocol stacks anywhere from the top to the bottom of the stack. A TCP-level gateway allows for two versions of TCP, one for the wired and another for the wireless network, on top of which a secure TLS channel can be established all the way from the mobile device to the server (see Fig. 9.b). The availability of a wireless profile of the TLS protocol, which includes cipher suites, certificate formats, signing algorithms, and the use of session resume, enables end-to-end security support at the transport level allowing interoperability for secure transactions.

5.3 Security visibility and configurability

Although the security measures provided by the SN should be transparent to the end user, visibility of the security operations as well as the supported security features should be provided. This may include: a) indication of access network encryption; b) indication of network wide encryption; and c) indication of the level of security (e.g., when a user moves from 3G to 2G).

Configurability enables the mobile user and the HE to configure whether a service provision should depend on the activation of certain security features. A service can only be used when all the relevant security features are in operation. The configurability features that are suggested include: a) enabling/disabling user-USIM authentication for certain services; b) accepting/rejecting incoming non-ciphered calls; c) setting up or not setting up non-ciphered calls; and d) accepting/rejecting the use of certain ciphering algorithms.

5.4 Network-wide user data confidentiality

Network-wide confidentiality is an option that provides a protected mode of transmission of user data across the entire network. It protects data against eavesdropping on every link within the network, and not only on the vulnerable radio links. Whenever network-wide confidentiality is applied, access link confidentiality on user data between the MS and the RNC is disabled to avoid replication. However, access link confidentiality for signaling information, as well as user identity confidentiality is retained to facilitate the establishment of the encryption process. In Fig. 10, the network-wide encryption deployment is depicted.

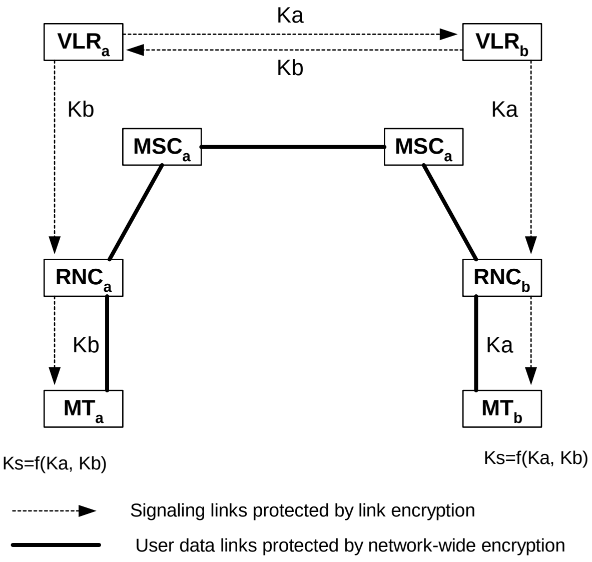
Ka
VLR$_a$
VLR$_b$
Kb
Kb
Ka
MSC$_a$
MSC$_a$
RNC$_a$
RNC$_b$
Kb
Ka
MT$_a$
MT$_b$
Ks=f(Ka, Kb)
Ks=f(Ka, Kb)
Signaling links protected by link encryption
User data links protected by network-wide encryption

Fig. 10: Network-wide encryption deployment

Network-wide confidentiality uses a synchronous stream cipher algorithm similar to that employed in the access link encryption. Initially, a data channel is established between the communicating peers indicating also the intention for network-wide encryption. VLRa and VLRb exchange cipher keys (Ka and Kb) for users a and b, respectively, using cross boundaries signaling protection, and then, pass them to the MSs over protected signaling channels. When each MS has received the other party’s key, the end-to-end session key, Ks, is calculated as a function of Ka and Kb. Alternatively, VLRs can mutually agree on the Ks using an appropriate key agreement protocol. Both key management schemes satisfy the lawful interception requirement, since Ks can be generated by the VLRs.

6 3G-security architecture evaluation

The existing UMTS security architecture provides advanced security services, and addresses many security concerns that have been listed in the context of next generation mobile networks. However, there are critical points, which need further elaboration and improvements. In the following, some security weaknesses that may cause network and service vulnerability are identified, as well as evolution perspectives that aim to enhance the level of security services and can be easily integrated in the 3G-security architecture are outlined.

6.1 Weaknesses

The mobile user identity and location is valuable information that requires protection. A possible weakness in 3G-security architecture is the backup procedure for TMSI reallocation [22]. Specifically, whenever the SN/VLR cannot associate the TMSI with the International Mobile Subscribers Identity (IMSI) because of TMSI corruption or database failure, the VLR should request the user to identify itself by means of IMSI on the radio path. Furthermore, when the user roams, and the SN/VLRn cannot contact the previous VLRo, or cannot retrieve the user identity; the SN/VLRn should also request the user to identify itself by means of IMSI on the radio path [13]. This may lead an active attacker to pretend to be a new SN to which the user has to reveal his permanent identity. In both cases, the IMSI, which represents the permanent user identity, is conveyed in clear-text on the radio interface violating user identity confidentiality.

Another critical point is that the users may be identified by means of the IMSI in signaling conversations in the wireline path. For example, the SN/VLR may use the IMSI to request the authentication data for a single user from its HE. Thus, user identity confidentiality and user location privacy rely on the security of the wireline signaling connections. NDS features protect signaling exchange in the wireline network architecture with IP and SS7 technologies, but these features are considered for the later versions of the UMTS standardization process, leaving the first one (R99) unprotected.

Firewalls were originally conceived to address security issues for fixed networks, and, thus, are not seamlessly applicable in mobile scenarios. They attempt to protect the clear-text transmitted data in the UMTS backbone from external attacks, but they are inadequate against attacks that originate from other mobile network malicious subscribers, as well as from network operator personnel, or any other third party that gets access to the UMTS core network. Mobility may imply roaming between networks and operators, possibly changing the source address, which because of the static configuration of firewalls, may potentially lead to discontinuity of service connectivity for the mobile user. Moreover, the firewalls security value is limited because they allow direct connection to ports and cannot distinguish services.

Similarly, the current type of VPN fails to provide the necessary flexibility to establish reliable secure connections for typical mobile users. VPN services for UMTS subscribers can be established in a static manner between the border gateway of the UMTS core network and a remote corporate security gateway. This makes the realization of VPN services feasible only between the security gateway of a large organization and a mobile operator, when a considerable amount of traffic requires protection. Thus, if the static VPN parameters or the VPN topology has to be changed, then the network administrators in both ends must reconfigure it. Furthermore, the aforementioned security scheme can provide VPN service neither to individual mobile users, that may require on demand VPN establishment, nor to enterprise users that may roam internationally.

In case that a mobile user uses the WAP architecture (ver. 1.2.1), data privacy is not guaranteed. Although encryption is used, the WAP gateway constitutes a security hole since, inside the gateway data are transmitted un-encrypted. WTLS is only used between the mobile device and the gateway, while TLS can be used between the gateway and the web server. From a security point of view, the gateway should be considered as an entity-in-the-middle. This means that data exchanged may be available to people with privileged access to the WAP gateway, and, thus, the privacy of the data depends the gateway's internal security policy.

WAP 2.0 does address the "gap" in security caused by protocol translation at the WAP gateway. However, the mobile phone would have to use an IP protocol stack, at the expense of larger latency and bandwidth consumption. Although TLS can be used to secure the communication of any application, it must be integrated into the application, and, thus, to a large extent it is used for web-based applications. Interaction with the end-user is needed, for example, to check with whom a secure session has been established, or to explicitly request the client to authenticate with the server. TLS is generally a resource consuming protocol for deployment in mobile devices with limited processing capabilities, and low bandwidth/high latency wireless networks. Moreover, the operation overhead may be increased by complex key-exchange procedures in case that the protected service contains cross-references to other services.

The network-wide encryption may also encounter problems when transcoding is used. Voice calls may need to be transcoded when they cross network borders, meaning that voice data may have to undergo change, such as, bit rate change or some other transformation. It is not possible to apply such transformation on an encrypted signal, which implies that the signal has to be decrypted before transcoding. Furthermore, the network-wide confidentiality lacks flexibility, and it is not applicable to all types of service in different mobile scenarios. Specifically, it is limited to protecting the communication between mobile subscribers.

Consequently, there is a lack of a general-purpose mechanism that can provide advanced security services to user data traffic according to the particular end-user needs, inside and outside the UMTS core network. Firewall technology cannot adequately ensure data transfer within the UMTS core network. Static VPN deployment and the network wide confidentiality option cannot be applied to all mobile scenarios protecting all type of potential services. Moreover, the WAP architecture and application layer security are mainly related to web-based applications, which can integrate TLS or WTLS functionality. Next generation mobile subscribers require dynamic, flexible, client initiated security mechanisms, which will be available anywhere – anytime. They should provide customized security services to data traffic, and guarantee interworking with existing and forthcoming network infrastructure, taking into account the end-user mobility and the mobile network characteristics.

6.2 Proposed improvements

The weak points in the 3G-security architecture, identified in section 6.1, may lead to compromises of end-user or mobile network security. In this section, enhancements that aim to improve the 3G-security architecture, as well as to provide advanced security services to end-user data traffic within and outside the UMTS core network are outlined. The proposed enhancements can be easily integrated in the existing network infrastructure and operate transparently to UMTS network functionality.

To prevent the exposure to threats of the permanent mobile user identity, where the clear-text IMSI paging procedure is being used, the employment of two additional temporary user identities has been proposed. Specifically, when the SN/VLR has failed to page a mobile user using the current TMSI, it can try to page him using an alternative temporary identity that also resides in the VLR, and thus, the use of IMSI can be avoided. If none of the TMSI is valid, or both TMSI are corrupted, the user is not attached to the network.

In case of a VLR database recovery, or a corruption of the TMSI in the VLR, the SN/VLR requires a second temporary identity by which it can page the user. This temporary identity has to be provided by the user's HE, otherwise it cannot be assured that it is available at the SN/VLR after a database recovery. For similar reasons, it cannot be provided to the SN/VLR in advance. The TMSI_{HE} should have a limited lifetime to prevent potential intruders to link it to the permanent user's identity [9].

NDS features protect signaling exchange in the wireline network architecture over IP and SS7 technologies. However, these features are considered for the later versions of the UMTS standardization process, leaving the first one unprotected. UMTS R99 is the precursor of 3G networks and needs to prove the necessity of transition from 2G+ to 3G. Apart from the higher access rate and the advanced QoS features, another reason that should enforce this transition would be the enhanced security services that 3G systems offer to the involved parties. Therefore, it is necessary that signaling exchange protection in the wireline network architecture be applied to the entire set of the UMTS releases.

Another issue that will empower the security services being provided in 3G networks is the advanced protection of user data traffic, either globally (end-to-end) or within the wireline network, according to the end-user needs. Complementary to the network wide confidentiality option and the WAP security, the incorporation of VPN technology in the 3G-security framework will further increase the supported level of protection by providing general-purpose security services at the network layer. On demand, customized VPN services are well suited to mobile users, which require anywhere – anytime connectivity. Moreover, VPN technology guarantees interworking with existing and forthcoming IP terrestrial network infrastructure.

The end-to-end VPN scheme [10] enhances mobile users' privacy in both the radio path and wireline network. The VPN functionality, which is based on the IPsec protocol suite, is integrated in the communicating peers, and thus, data traffic remains encrypted for the entire route between them. It has minimal impact on the existing network infrastructure, and provides the best security services to the end users. On the other hand, the main drawback of this scheme derives from the fact that each MS must have the appropriate software (IPsec) in order to apply the required security policy. This imposes computational costs on the lightweight end-user devices, and duplicates encryption (packet encapsulation) over the expensive radio interface. Moreover, the end-user must be aware of when encryption is required.

An alternative approach to the end-to-end VPN scheme pertains to a network-based scheme [11, 12]. This scheme integrates VPN functionality into the network infrastructure, and eliminates the need for end-user involvement. The network operator offers responsive, reliable, and flexible VPN services, so that the administrative and the computational overheads for the end-user are minimized. By relying on a sequence of concatenated protection mechanisms (UTRAN ciphering and VPN deployment), it is possible to provide secure data transfer without requiring an extra tunnel overhead on the radio link, or the implementation of computationally intense encryption algorithms in the MS. The VPN functionality is also based on IPsec. For VPN initialization and key agreement procedures an Internet Key Exchange (IKE) protocol proxy scheme has been proposed, which enables the mobile user to initiate a VPN, while shifting complex key negotiation to the network infrastructure. The deployed VPNs provide maximal security services to end users, operate transparently to the MSs movement, and are compatible with the legal interception option. The required enhancements for security service provision can be integrated in the existing network infrastructure, and therefore, the security scheme can be used as an add-on feature of the UMTS.

There is significant interest in such solutions both by customers, seeking to reduce support costs, and by network operators, seeking new revenue sources. By placing security functionality in the UMTS access network or in the UMTS border, the network-wide or the border-based VPN scheme are deployed, respectively [11, 12]. In the network-wide scheme the deployed VPN is extended over the UMTS backbone and the public Internet, while in the border-based it expands only in the public Internet segment. The particular scheme that will be selected in a potential deployment scenario depends on the required level of security services, as well as the network topology.

7. Conclusions

The evolution of 3G-networks signifies a shift towards open and easily accessible network architectures, which raises major security concerns. Thus, all stakeholders are interested in the security level supported in 3G-systems because of the identified risk that prospective users might perceive 3G-networks as not trustworthy, and not use them. This paper has elaborated on the security framework in 3G mobile networks. The security requirements imposed by the different types of traffic, and by the different players involved (mobile users, serving network and service providers) have been outlined. The security architecture, which comprises all the security mechanisms that are projected for the next generation mobile networks, has been presented and analyzed. 3G-security attempts to ensure that all information generated by or relating to a user, as well as the resources and services provided by the serving network and the home environment, are adequately protected against misuse or misappropriation. The employment of traditional security technologies, originally designed for fixed networking, such as firewalls, and static VPN, in order to safeguard the UMTS core network from external attacks, as well as to protect user data when conveyed over the public Internet have been examined. Critical points in the 3G-security architecture that may cause network and service vulnerability have been identified. Finally, proposals that aim to enhance the 3G-security architecture, as well as to provide advanced security services to end-user data traffic within and outside the UMTS core network have been briefly discussed. The proposed enhancements can be easily integrated in the existing network infrastructure, and operate transparently to UMTS network functionality.

Acknowledgments

The authors would like to thank Nikos Loukas and the anonymous reviewers for their comments and suggestions.

References

[1] GSM 03.60, “GPRS, Service Description, Stage 2”, 1998.
[2] 3GPP TS 23.002 (v3.6.0) “Network Architecture”, release ’99, Sept. 2002.
[3] 3GPP TS 25.401 (v3.10.0) “UTRAN Overall Description”, release ’99, Sept. 2002.
[4] 3GPP TS 23.002 (v4.5.0) “Network Architecture”, release 4, Sept. 2002.
[5] 3GPP TS 23.002 (v5.8.0) “Network Architecture”, release 5, Sept. 2002
[6] S. Kent and R. Atkinson, “Security Architecture for the Internet Protocol,” RFC 2401, Nov. 1998.
[7] 3GPP TS 21.133 (v3.2.0) “3G Security, Security Threats and Requirements”, release ’99, Dec. 2001.
[8] 3GPP TR 33.900 (v1.2.0) “A Guide to 3G Security”, Jan. 2000.
[9] “USECA UMTS Security Architecture” AC336/ATEA/WP23/DS/P/08/1, USECA project, Deliverable 08, May ’02
[10] C. Xenakis, E. Gazis and L. Merakos, “Secure VPN Deployment in GPRS Mobile Network,” Proc. European Wireless 2002, Florence Italy, Feb. 2002, pp. 293-300.
[11] C. Xenakis and L. Merakos, “Dynamic Network-based Secure VPN Deployment in GPRS,” Proc. PIMRC 2002, Lisboa, Portugal, Sept. 2002, pp. 1260-1266.
[12] C. Xenakis and L. Merakos, “On Demand Network-wide VPN Deployment in GPRS,” IEEE Network, Vol. 16, No. 6, Nov/Dec. 2002, pp. 28-37.
[13] 3GPP TS 33.102 (v3.12.0) “3G Security, Security Architecture”, release ‘99, June 2002.
[14] 3GPP TS 22.100 (v3.7.0) “UMTS phase 1 Release 99”, release ’99, Oct 2001.
[15] 3GPP TR 33.908 (v3.0.0) “3G Security; General report on the Design, Specification and Evaluation of 3GPP Standards Confidentiality and Integrity Algorithms”, release ’99, March 2000.
[16] TS 29.002 (v3.12.0) "Digital cellular telecommunications system (Phase 2+); Mobile Application Part (MAP) specification", release ’99, March 2002.
[17] 3GPP TS 33.210 (v5.1.0) “3G Security; Network Domain Security: IP network layer security”, release 5, June 2002.
[18] 3GPP TS 33.200 (v4.3.0) “3G Security; Network Domain Security; MAP application layer security”, release 4, March 2002.
[19] D. Harkins and D. Carrel, “The Internet Key Exchange (IKE)”, RFC 2409, Nov. 1998.
[20] 3GPP TS 31.111 (v3.7.0) “USIM Application Toolkit (USAT)”, release ‘99, Dec. 2001.

[21] B. Gleeson, A. Lin, J. Heinanen, G. Armitage, and A. Malis, “A Framework for IP Based Virtual Private Networks”, RFC 2764, Feb. 2000.

[22] 3GPP TS 24.008 (v3.13.0) “Mobile Radio Interface Signaling Layer 3 specification; Core Network Protocols – Stage 3”, release ‘99, Sept. 2002.

[23] 3GPP TS 35.205 (v3.0.0) “3G Security; Specification of the MILENAGE Set: An example algorithm set for the 3GPP authentication and key generation functions f1, f1*, f2, f3, f4, f5, and f5*”, release ‘99, April 2001.

[24] Wireless Application Forum (WAP), WAP specifications,
URL: http://www.wapforum.org/what/technical.htm.

Abbreviations and symbols

2G	Second Generation
3G	Third Generation
AK	Anonymity Key used in 3G
AMF	Authentication management field
AN	Access Network
AUTN	Authentication Token
AuC	Authentication Centre
AV	Authentication Vector
BTS	Base Transceiver Station
BSC	Base Station Controller
BSS	Base Station Subsystem
CK	Cipher Key used in 3G
CS	Circuit Switched
EIR	Equipment Identity Register
ESP	Encapsulation Security Payload
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Services
GSM	Global System for Mobile communication
GTP	GPRS Tunneling Protocol
HE	Home Environment
HLR	Home Location Register
HTTP	HyperText Transfer Protocol
IK	Integrity Key used in 3G
IKE	Internet Key Exchange
IMSI	International Mobile Subscriber Identity
IP	Internet Protocol
IPsec	IP security
KAC	Key Administration Center
MAC	Message Authentication Code
MAP	Mobile Application Part
MAPsec	MAP security
MS	Mobile Station
MT	Mobile Terminal
MSC	Mobile Switching Centre
NE	Network Entities
NDS	Network Domain Security
OSA	Open Service Architecture
PS	Packet Switched
Rel-4	Release 4
Rel-5	Release 5

R99	Release ‘99
RAND	Random challenge
RES	(expected) user response to challenge in GSM/GPRS
RNC	Radio Network Controller
SA	Security Association
SG	Security Gateways
SRES	Signed response
SN	Serving Network
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
SG	Security Gateways
SP	Service Provider
SS7	Signaling System 7
TE	Terminal Equipment
TCP	Transmission Control Protocol
TMSI	Temporary Mobile Subscriber Identity
TLS	Transport Layer Security
UICC	UMTS Integrated Circuit Card
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestrial Radio Access Network
USIM	User Services Identity Module
VLR	Visitor Location Register
VPN	Virtual Private Network
WAE	Wireless Application Environment
WAP	Wireless Application Protocol
WCDMA	Wideband Code Division Multiple Access
WDP	Wireless Datagram Protocol
WSP	Wireless Session Protocol
WTLS	Wireless Transport Layer Security
WTP	Wireless Transport Protocol
XRES	Expected Response
XMAC	Expected Message Authentication Code
||	String concatenation
⊕	Exclusive or
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
K	Long-term secret key shared between the USIM and the AuC in 3G
Ks	Session secret key
A	Interface between an MSC and an BSS
Abis	Interface between a BTS and a BSC
Gb	Interface between an SGSN and a BSS
IuCS	Circuit-Switched interface between an RNC and a core network
IuPS	Packet-Switched interface between an RNC and a core network
Iur	Logical interface between two RNCs
Iubis	Interface between an RNC and a Node B

Um	Radio interface between a mobile station and a GSM fixed network part.
Uu	Radio interface between UTRAN and a User Equipment