INTRODUCING PKI TO ENHANCE SECURITY IN FUTURE MOBILE NETWORKS

Georgios Kambourakis, Angelos Rouskas and Stefanos Gritzalis
Department of Information and Communication Systems Engineering,University of the Aegean, Samos 83200, Greece

Abstract Current wireless network standards perform user authentication, signaling and data encryption, as well as message integrity protection, by utilizing only symmetric key methods. However, as mobile networks are evolving into full-IP and the communication is envisaged to change from second generation (2G) person-to-person model to fourth generation (4G) machine-to-machine model, there is greater demand to provide more flexible, reconfigurable and scalable security mechanisms that can advance in a many-to-many trust relationship model. Employing public key methods in many-to-many schemes drops the requirement for a secure channel to transfer keys between two communication parties, thus providing the appropriate scalability to the whole system. With a large number of different network technologies and operators, expected in the future mobile communications environment, that should frequently and seamlessly interwork with each other, and a constantly increasing population of communication parties, capturing the full benefits of open channel key transfers and scaling public key methods requires Public Key Infrastructure (PKI). In this paper, we discuss and investigate different ways to take advantage of a proposed PKI system. From the network side, we investigate how PKI can provide future inter/intra mobile core network security, while from the user's perspective we present solutions that far enhance authentication procedures and end-to-end communication model trust. We show that PKI offers the appropriate framework to overcome symmetric key based security inefficiencies, providing powerful solutions to protect both network core signaling and user's data from potential intruders.

Keywords: PKI; Mobile Networks; IKE; Network Domain Security; SSL/TLS; IPsec

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35691-4_52

D. Gritzalis et al. (eds.), *Security and Privacy in the Age of Uncertainty*

1. INTRODUCTION

An identified security weakness in 2G systems is the absence of security in the core network. Originally, this was not a problem, since 2G networks were closed networks with very little interworking among different 2G operators and between 2G operators and the Internet. Nevertheless, in a future wireless communication environment, like 3G and beyond, that will require frequent interworking of many different network technologies and providers, there will also be a greater need for advanced security protection. Moreover, the introduction of IP, used not only for signaling traffic, but also for user traffic, as the network layer in the GPRS backbone network and later in the UMTS network domain, raises further reasons to worry about. The introduction of IP states not only a shift towards packet switching, but also a shift towards completely open and easily accessible networks. From a security point of view, a whole new set of threats and risks must be faced. For example, in next generation systems the protection of the core network signaling protocols will be a clear and essential requirement.

The adaptation of Public Key Infrastructure elements in future wireless networks will substitute long-term symmetric key relationships, with a flexible, reconfigurable and scalable public key based mechanism. This will not only provide the appropriate level of inter/intra operator trust, but it will also offer solutions that far enhance user-to-network confidence and end-to-end security options.

The rest of this paper is organized as follows. In Section 2, we provide an overview of the current 3G-inter/intra security options and explain how PKI can adapt to existing architecture. In Section 3, we propose and analyze some solutions that can be implemented, to provide inter/intra operator trust, user-to-network and end-to-end security by taking advantage of PKI. Finally, the paper is concluded in Section 4.

2. EXISTING 3G CORE NETWORK SECURITY AND PKI

2.1 Overview of 3G Inter/Intra network Security

Global Mobile System (GSM) and Universal Mobile Telecommunication System (UMTS) networks (Figure 1), use Mobile Application Part (MAP) protocol for the exchange of signaling messages between network Elements (NEs). User profile exchange, authentication, and mobility management are performed using MAP. MAP runs typically over the Signaling System Number 7 (SS7) protocol stack.

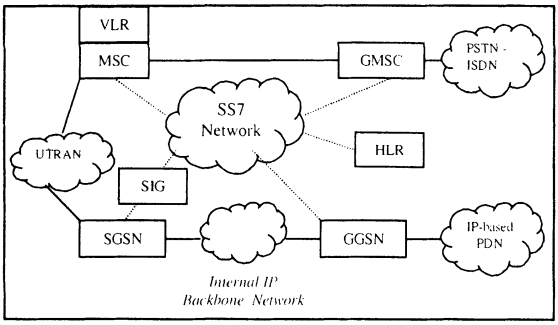


Figure 1. UMTS Architecture

3GPP has also defined a mechanism for protecting the MAP protocol at the application layer [1], [2]. MAP may also be protected at the network layer when IP is used as the transport protocol. However, when internetworking with networks using SS7-based transport is necessary, protection at the application layer shall be used. For this reason a new protocol header has been developed to protect MAP messages, much in the same way as the Encapsulating Security Payload (ESP) protocol protects IP packets. This new protocol is called MAPsec. While MAP runs over SS7, MAPsec and Internet Key Exchange (IKE) always run over IP (Figure 2). Therefore, it is assumed that nodes implementing MAPsec always have IP connectivity in addition to SS7 connectivity [2]. In the 3GPP architecture MAPsec is typically running between two different network operators, and the same Security Associations (SAs) are shared by a number of NEs. The necessary MAPsec-SAs between networks are negotiated between the respective Key Administration Centers (KACs) of the networks.

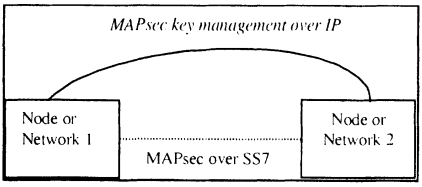


Figure 2. Architecture for MAPsec Security

On the other hand, for native IP protocols, as in the GPRS backbone network [3], security shall be provided at the network layer. The security protocols to be used are the IETF defined IPsec suite [4]. The UMTS network domain control plane is sectioned into security domains, which typically coincide with operator borders. The borders between the security domains are protected by Security Gateways (SEGs) as shown in Figure 3.

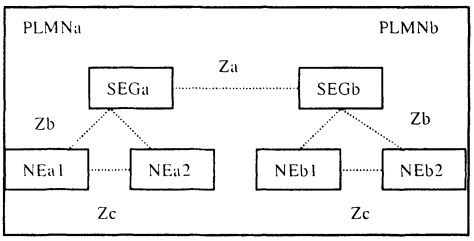


Figure 3. Network Domain architecture for IP-based protocols. (Dashed lines represent IKE connections)

All network domain traffic shall pass though a SEG before entering or leaving the security domain. Taking that into account, Network Domain Security (NDS)/IP will only support tunnel mode IPsec SAs, ESP and main mode. SEGs shall offer capabilities for secure storage of long term keys used for IKE authentication, so NDS/IP will only support Internet Security Association and Key Management Protocol (ISAKMP) SAs with pre-shared keys [5].

Only the inter-security domain SA IKE negotiations over the Za interface shall be mandatory, while the Zb interface is optional. Concluding, there is normally no NE-to-NE interface for NE belonging to different security domains.

2.2 Adding PKI to mobile networks

As we already mentioned, proposals and technical specifications for core network security in 3G, are based on IPsec. Agreements on keys and security associations are carried out on a bilateral basis between operators. Taking into account that the number of network elements of each operator increases and that the interworking between a high number of networks of different technologies will be more intense, a more scalable solution would be to replace those relationships with a PKI [6], [7]. So secure communications can be achieved without having to generate and distribute long-term secret keys.

Comparing an asymmetric key system with a symmetric one, we note the following:

- The number of keys needed in a symmetric key system with n network elements communicating with each other is $O(n^2)$. On the other hand, in a public cryptosystem, the corresponding need for keys is $O(n)$. Therefore, when n increases, the costs in terms of key generation and distribution associated with the introduction of a new network element are quite different. In the symmetric model, we need to establish n new secret keys, while in the asymmetric case we only need 2 new keys (private + public) for any new network element.

- Pre-shared secrets are a rather inflexible way to provide authentication. A properly designed PKI, which supports digital certificates, will offer more dynamic, flexible and scalable mechanisms to issue certificates for new network elements and to revoke certificates that are no longer valid.
- One basic requirement and assumption in both GSM and UMTS, is that the Home Environment (HE) has to trust the Serving Network (SN), e.g. for the Authentication and Key Agreement (AKA) procedure. However, in future systems, where many different technologies, owned by different network operators, must frequently and seamlessly interwork, this is no longer the case. By introducing a Trusted Third Party (TTP) the requirement for bilateral trust is reduced.
- PKI can be used for authentication and symmetric key encapsulation and transport procedures, while derived symmetric session keys can be used to support confidentiality. Thus, we can by-pass the known public key cryptosystem disadvantages of key lengths and computational load.
- From the user scope, the implementation of public key algorithms in Mobile Stations (MSs) had been considered to be resource demanding. However, the increased processing requirements of IP capable terminals have driven towards high power computational platforms, which are now becoming ordinary in wireless devices. In addition, advanced standards such MExE and WAP have also moved forward to introduce public key methods. This development strengthens the assertion that PKI has become an acknowledged component of standards that deal with many-to-many, complex relationships.

Furthermore, as IP-based networks are introduced to serve a large variety of applications, that may involve many and different network/service operators, complex and flexible communication relationships are necessary, which in turn demand a complex trust model. In many cases, the communication parties may not have pre-arranged security agreements. So, in order unknown partners perform mutual authentication and establish session keys, a public key based digital signature that is supported by a PKI will satisfy security needs. For example, a Session Initiation Protocol (SIP) registration server, either proxy or redirect [3], may not share any symmetric key with the User Equipment (UE). Instead, a digital signature may be an appropriate way to authenticate the proxy server.

Certainly, the support of asymmetric key services by a wireless network requires the adaptation of some PKI elements, which are not necessarily part of the speculated network core. Figure 4 depicts the necessary PKI elements that should be included in the UMTS architecture.

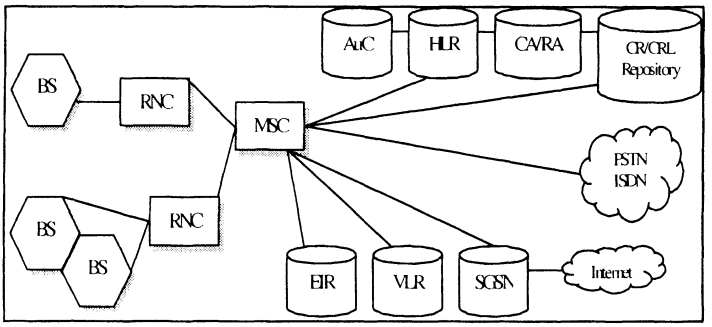


Figure 4. UMTS architecture and PKI

More specifically, we assume the following:

- There is some sort of Certification Authority (CA/RA) per Public Land Mobile Network (PLMN) operator, which issues and revokes certificates. A Registration Authority (RA) can offload the CA with certain functions i.e. establish and confirm identities of a new network element, generate keying material, perform certain key/certificate life cycle functions etc.
- There is one-at-least digital certificate database, which stores all the digital certificates and is being managed by the PLMN's CA.
- There is one at least revoked certificates database (CRL-database), which is being managed by the PLMN's CA and is accessible from all network elements that belong to the mobile network core.
- Web servers or FTP servers can be used to store certificates and CRLs. Certificate revocation can be periodic or Online Certificate Status Protocol (OCSP) based [8].
- CAs, which belongs to different PLMN's issue (off-line) cross-reference certificates for inter-PLMN trust relationships [8]. For example in the case of two PLMNs with the corresponding certification authorities CA_a & CA_b, CA_a issues Cert(CA_a)CA_a[1] (the root certificate) and Cert(CA_b)CA_a. Respectively, CA_b issues Cert(CA_b)CA_b and Cert(CA_a)CA_b.
- Cross-Reference certificates are cached in local Security Gateways SEGs (which probably implement firewall policies among other things) on the borders of IP security domains. Every PLMN can use one or more SEG, in order to balance inter network traffic.
- Every network element possesses a key pair (private + public), and the corresponding digital certificate (intra-operator trust). NE's private key and the public key of the local CA are stored locally in a secure manner.
- If we are planning to extend PKI usage to the user, primarily for authentication and symmetric key encapsulation, we can assume the following:

[1] Cert(X)y = Public key certificate of X with format X.509v3 (or subset) issued by Y.

- The USIM-4G smart card should be a crypto-card with good pseudo-random (or random) generation capabilities and in-built crypto accelerator chip.
- Every subscriber possesses a key pair and his private key is stored in his USIM-4G card. The keys are associated with the user at registration time.
- Furthermore, the USIM-4G card is pre-loaded with all the CA's public keys, which exist in the particular PLMN.

3. PROPOSED PKI SOLUTIONS

3.1 PKI-based Intra/Inter Future Network Domain Security

With network domain security we mainly mean secure communications between network elements. Thus, by introducing a PKI to a future wireless network we can use powerful protocols to protect signaling and user traffic both between inter-network and intra network elements.

Three connections have to be protected as shown in Figure 3:

1. Za or SEG-to-SEG (inter-operator security),
2. Zb or SEG-to-NE (intra-operator security) and
3. Zc or NE-to-NE (intra-operator security).

One candidate for this task is IPsec [9], and IKE in particular. As we already mentioned, 3GPP currently uses pre-shared secrets for IKE phase I. This means that each NE has to be configured with a password that is associated with the remote system's IP address being authenticated. Note, however, that the keys to be used for encryption and authentication (SKEYID_*), after the completion of phase I, have been generated solely based on the peer's IP address [10]. So, in scenarios where the IP address is dynamic, the responder cannot maintain pre-shared secrets indexed by an IP address that may not be known at that time. Remote access solutions are an example where the initiator's IP address may be different for each connection (road-warrior cases) [9]. Additionally, the main drawback in pre-shared secret key authentication is the lack of a secure and scalable mechanism for exchanging pre-shared secret keys. That is appropriate only in a rather small-scale environment with a restrained number of systems, in which the set of peers is known in advance. However, if a pre-shared secret key is compromised, there is no universal method to alert the peer and launch a replacement.

An alternative solution based on PKI, can overcome these shortcomings. According to this, IKE is used for key exchange over the Za, Zb and Zc interfaces, while the authentication could be based on digital signatures with certificates instead of pre-shared secrets. The sequence of messages for this procedure is shown in Figure 5.

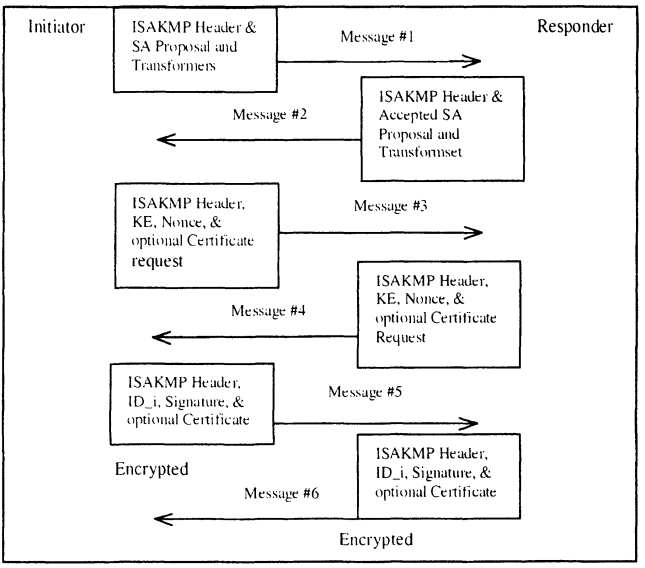


Figure 5. Main mode with digital signature

We note that, in such a case, the keys to be used for encryption and message authentication are generated based solely on the peer's nonce and Diffie-Hellman key value (SKEID = prf[2] (Nonce_i | Nonce_r, DH_Key). For system authentication, a certificate request can be included to obtain the public key of the peer if the initiator does not already have it. The peer must have the other's public key to validate the signature and authenticate the peer in the third exchange (messages 5 & 6). Also, the use of certificates in such a scheme provides for non-repudiation [8], [10].

Another solution, which benefits by the incorporation of PKI, is the use of SSL/TLS to protect communications between security gateways and probably between NEs. SSL/TLS is a flexible, session-oriented protocol that provides security at the transport layer, a higher layer in the TCP/IP stack than IP [11]. SSL/TLS has many of the advantages of IPsec and the successful introduction of the protocol in the wired Internet has proved its usability and effectiveness. Likewise, SSL/TLS can be part of an all-IP mobile environment. In this context, we provide below a short description of

[2] Pseudo random function

the necessary SSL/TLS handshake protocol message exchanges between SEGa and SEGb, to protect the borders of different operators.

1. SEGa, acting as a client, initiates the connection with a Hello message, which contains SEGa's security options and a session ID.
2. SEGb, acting as a server, replies with its Hello message. If SEGb support cryptographic and compression methods common to SEGa, those are included in its message. Otherwise, the connection is terminated. In addition, SEGb sends its 32-bit random number and a session ID. If the latter is equal to SEGa's session ID it is implied that the parties are going to use security parameters agreed on a previous session (this is the session resumption option, which considerably speeds-up the overall process [11]. Otherwise, SEGb generates a fresh session ID number denoting a new connection.
3. SEGb sends its digital certificate to SEGa, along with the appropriate cross-certificate, in a certificate chain and requests SEGa's certificate, concluding its part of negotiation with a HelloDone Message.
4. Once SEGb's certificate and cross-certificate have been validated, SEGa generates a pre-master secret, encrypts it using SEGb's public key and sends the encapsulated key to SEGb.
5. SEGb validates SEGa's certificate chain and decrypts the pre-master secret by using its private key.
6. Both parties convert the pre-master secret into master secret. The master key will be used for ciphering and MAC computations.
7. SEGa and SEGb send CipherSpec + Finished messages to each other. Note that the finished messages are cryptographically and integrity protected making use of previously negotiated parameters.

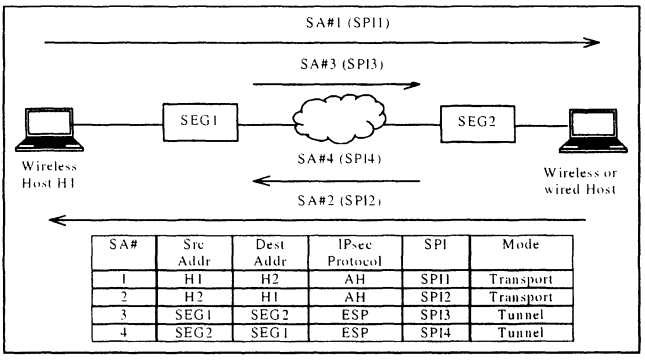
3.2 PKI and wireless network user benefits

From the user's side, a PKI can support the appropriate reconfigurable infrastructure, which offers great flexibility and scalability in an all-IP wireless environment. In this way, we can provide for authentication and end-to-end security solutions, which far enhance the user's trust, in a continuously evolving environment.

It is still a common misbelief, that mobile devices are not ready for "hungry", in terms of memory and processing power, public key computations. However, that is partially true, since contemporary wireless devices are featuring advanced architectures with StrongArm processors up to 206MHz, memory capacities of 64MB RAM and 32MB ROM, support for java applications and strong operating systems which can support a variety of applications and protocols. Besides that, these trends has also driven smart cards toward more advanced architectures, all the way to where we are beginning to see 32-bit RISC-based ARM processors in smart cards.

These cards based on such modern chips from companies like Atmel and Infineon are just appearing in the market, and they can effectively store and protect the subscriber's private key, generate good pseudo-random values and take over of symmetric key (un)wrapping functions [8]. The mobile's device processor can efficiently carry out the rest of the calculations, needed by protocols like IPsec and SSL/TLS. Last but not least, wireless network speed is continuously increasing, thus offering the necessary bandwidth.

In an IP-enabled mobile device with the aforesaid characteristics, IPsec, can effectively secure authentication and user traffic, therefore, providing a secure end-to-end channel (Figure 6). Once again, IKE with authentication based on digital certificates will be used instead of pre-shared secrets. Road-warrior cases can also be effectively authenticated using this scheme.



SA#	Src Addr	Dest Addr	IPsec Protocol	SPI	Mode
1	H1	H2	AH	SPI1	Transport
2	H2	H1	AH	SPI2	Transport
3	SEG1	SEG2	ESP	SPI3	Tunnel
4	SEG2	SEG1	ESP	SPI4	Tunnel

Figure 6. Nested headers for end-to-end IPsec protection

For example, consider the following scenario:

"A business employee has IPsec-based Virtual Private Network (VPN) client software installed on his laptop, which is connected to his wireless network provider, via his mobile phone. Also assume that the employee is roaming to a foreign (serving) network. When he connects, he is being authenticated by IKE sending its digital certificate and receiving Visitor Location Register's (VLR) certificate and cross-reference certificate. When IPsec's SAs have been created, VPN client filters the traffic, watching for IP packets destined to the employee's head office. It allows any traffic not going to the head office to pass unprotected. When however the client spots a packet that is addressed to the head office intercepts it. It then uses IPsec services to transmit the packet securely and to assure that all traffic back from the head office to him is also secure".

The first thing that the VPN client does is to establish a bi-directional IPsec Security Association (SA) with the head office server. IKE (ISAKMP) [12] defines the framework how the VPN client and server set up security associations. It does however require the use of digital signatures within the

authentication section. This means that the VPN client and server must have IPsec public key certificates to be able to establish a security association.

Taking into account the aforesaid technological trends, SSL/TLS can similarly provide for user authentication and end-to-end security. Until now performance considerations in using SSL/TLS in a resource-constrained environment drove wireless designers to choose a different, incompatible and insecure gateway oriented security protocol for their mobile clients, like in the case of WAP [13], [14]. Provided that it is possible to develop a usable, in terms of performance, implementation of SSL for a handheld device, we can employ a more secure, flexible and reconfigurable Authentication and Key Agreement (AKA) procedure for 4G systems [16]. The ASPeCT project [17] has demonstrated that public-key authentication is possible and GSM and UMTS applications can coexist on a single smart card. A recent study has also shown the feasibility of SSL/TLS in handheld wireless devices [18].

Moreover, the PLMN capability to pre-load an asymmetric key pair in the USIM-4G card combined with public key infrastructure, allows the CAs to publish temporary certificates (attributes or service certificates) [8], which will admit the user access to specific time-limited services and resources [19]. To be more specific, we present the following scenario:

"The user applies to the serving network to provide him a specific time-limit service and signs his request with his private key, which is stored in the USIM-4G. The request is forwarded to the local CA, which checks its validity and then publishes -on the fly- a temporary signed by it certificate, which specifically designates the service type and its expiration time. The certificate is forwarded back to the user, who can use it correspondingly in a SSL/TLS protected channel, communicating with the provider of the specific service. One of the advantages of these certificates having a short life, is that they will not usually need to be revoked and will therefore not need to be included in any CRL. They may also not require revocation if they are issued in respect of a pre-paid subscription service. This mechanism can also support non-repudiation services."

4. CONCLUSIONS

As users rush to adopt IP technology and want mobile access to IP networks, they also become aware of the need for security features and protection of their privacy. The constantly increasing population of users expects from wireless operators to provide features that will protect their data while in transit, safeguard their billing and customer information, and offer availability and quality comparable to that of the wired services. Thus, more flexible, dynamic and scalable mechanisms are necessary in order to

support on-demand services and all-IP end-to-end solutions in a many-to-many trust model integrated with the Internet environment. In this paper, we proposed several alternative procedures based on PKI infrastructure introduced in the mobile network architecture for providing future inter/intra mobile core network security, enhancing authentication procedures and end-to-end communication model trust. We showed that PKI offers the appropriate framework to overcome symmetric key based security inefficiencies, providing powerful solutions to protect both network core signaling and user's data from potential intruders.

5. REFERENCES

[1] 3GPP Technical Specification, MAP Application Layer Security, (TS 33.200 v. 5.0.0), March 2002

[2] J.Arko, R. Blom, "The MAP Security Domain of Interpretation for Internet Security Association and Key Management Protocol", <draft-arkko-map-doi-07.txt>, May 2002.

[3] Wisely, D., Eardley, P., & Burness, L., *IP for 3G*, Wiley, 2002.

[4] Kent, S. & Atkinson, R., *Security Architecture for the Internet Protocol*, RFC 2401, Nov. 1998.

[5] 3GPP Technical Specification, IP Network Layer Security, (TS 33.210 v.5.1.0), June 2002

[6] 3GPP TSG, "Using PKI to provide network domain security", Discussion Document S3-010622 SA WG3 Security – S3#21, Nov. 2001.

[7] 3GPP TSG, "Security Services using Public Key Cryptography", Discussion Document S3z000025 SA WG3 Security – S3#15bis, Nov. 2000.

[8] Duane, N. & Brink, J., *PKI Implementing and Managing E-Security*, Berkeley, RSA press, 2001

[9] Frankel, S., *Demystifying IPsec Puzzle*, Artech House, 2001.

[10] Tiller, J., A Technical Guide to IPsec Virtual Private Networks, Auerbach CRC Press, 2000.

[11] Thomas S., *SSL and TLS essentials*, New York, Wiley, 2000.

[12] Maughan, D., et al, Internet Security Association and Key Management Protocol (ISAKMP), RFC 2408, Nov. 1998.

[13] Khare R., "W* Effect Considered Harmful", IEEE Internet Computing, Vol. 3, no 4, pp.82-92, July/Aug. 1999.

[14] WAP forum WAP-217-WPKI, "Wireless Application Protocol Public Key Infrastructure Definition", April 2001, www.wapforum.org/what/technical.htm.

[15] 3GPP Technical Specification, 3G Security Architecture, (TS 33.102 v.4.3.0), December 2001.

[16] Kambourakis G., Rouskas A., & Gritzalis S., "Using SSL/TLS in Authentication and Key Agreement Procedures of Future Mobile Networks", In the Proc. of the 4th IEEE Int'l Conf. on Mobile and Wireless Comm. Networks. (MWCN), pp. 152-156, 2002.

[17] ASPeCT Project, *Securing the future of Mobile Communications*, http://www.esat.kuleuven.ac.be/cosic/aspect, 1999.

[18] Gupta V. & Gupta S., "Securing the Wireless Internet", IEEE Communications Magazine, Vol 39, no 4, pp. 69-74, Dec. 2001.

[19] 3GPP TSG, "Support of certificates in 3GPP security Architecture", Discussion Document S3-010353 SA WG3 Security – S3#19, July 2001.