Public Key Infrastructure (PKI) and Mobile Commerce
Issues, Implications and Future Trends

Nabeel A. Al-Qirim, United Arab Emirates University, UAE
Eman I. Al Haj Ali and Nurul I. Sarkar, Auckland University of Technology, NZ

INTRODUCTION

The use of mobile phones by individuals and businesses has grown in recent years from providing simple telephone and message services to the provision of fully-fledged and interactive mobile services and transactions on sophisticated PDAs and wireless-laptops. With this growing complexity in the use of mobile technology applications in the business environment, such as conducting commercial transactions (e.g. purchasing goods) and paying money online, there has been an increasing demand for securing and protecting sensitive/personal information, transactions and the corresponding electronic/mobile commerce technology infrastructure.

Transactions conducted through the Internet are usually between strangers and hence, lack the needed trust to complete such transactions. Specifically, there is a requirement to guarantee the secrecy (preserving privacy and confidentiality), authenticity (originality of users in the Internet) and authorisation, integrity (transaction modification), non-repudiation (enforceability of transactions and preventing double-spending) and necessity or denial of service (availability of hardware, software and services in general (web server)). The use of electro-mechanical devices, tools and procedures such as securing the hardware and software behind locked and protected rooms (e.g., access control devices), closed circuit camera TV (CCTV), voice, face, finger or hand or even eye/retina scanners could protect against many of these threats. However, these tools are useless when they face digital threats similar to the ones emerging from the Internet and external networks, e.g. threats from hackers, spoofers, snoopers, masquerades, etc., are just some of the examples. Let us not forget that most of these threats originate from legitimate employees within organizations, e.g. disgruntled/sacked employees – who have all the needed privileges to gain access to internal networks and create havoc, confusion, and disruption to the organization's systems and operations. These inherent problems in the wired electronic arena have been passed on to the wireless medium as well.

Encryption emerges as one possible tool to combat such digital threats at different levels of sophistication and success. This research discusses this tool in great detail and then focuses on the Public Key Infrastructure (PKI) encryption technique, as being the most stringent encryption tool available in the marketplace nowadays, and highlights the different issues facing PKI.

ENCRYPTION

Encryption is defined as a security control and a fundamental mechanism used to secure the content integrity of the transmitted information on the Internet (Busta, 2002; Minoli & Minoli, 1997). Encrypting the transmitted message so that unauthorised parties cannot intentionally or unintentionally reveal or/and tamper with the contents of the transmitted information can be achieved using different methods (Helms, Bushong, & Nelms, 2002):

- Secret or symmetric key (SK) encryption: is one-way of encrypting the contents of a message transmitted on the Internet using a secret key and an algorithm (e.g., RSA). The same secret key

is used to encrypt and decrypt the message. The integrity of the secret key encryption is only retained if users keep the secret key secret.

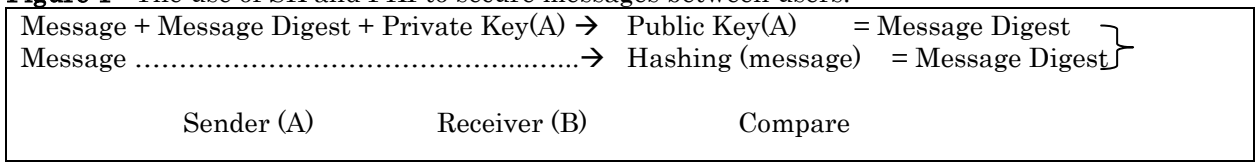
- Public key encryption: is another method of encrypting the contents of transmitted messages on the Internet using public/private keys (paired together). Electronic storefronts usually publish the public key on their website for consumers to encrypt the transmitted message and to keep the private key a secret. The paired public and private keys are numerically related to each other. The only person with the private key can decode the transmitted message encoded with the paired public key and vice-versa.

However, the above methods don't reveal for online shoppers whether they are dealing with real businesses or impersonators. The risk of having the transmitted information tampered with by unauthorised parties is still valid and hence, using other security standards on the web is highly recommended here to minimise such risks. To retain the integrity of the message during transmission on the Internet, this could be achieved through the use of Digital Signature and hashing techniques (Helms et al., 2002; King, 1997). This is explained next.

DIGITAL SIGNATURE

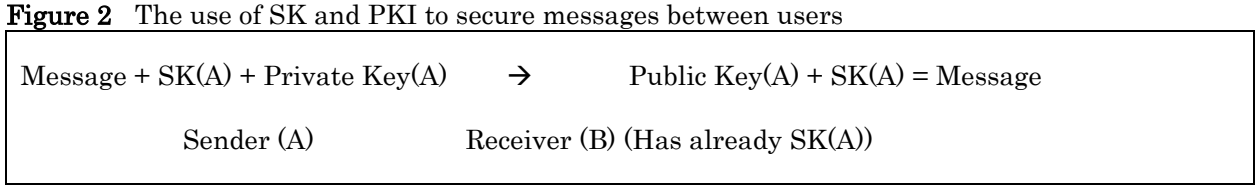
Digital signatures are simply created by passing the content of the transmitted message through a mathematical algorithm (hash function) (e.g., SHA-1, MD5) to produce a fixed size string called the message digest (or the hashed value) (King, 1997). The resulting hash value is then encrypted with the private key to produce the digital signature (Figure 1).

Figure 1 The use of SK and PKI to secure messages between users.



In some cases, users may use their secret key to further encrypt the digital signature which includes the original message and the hashed value and/or the public key so that the recipient with the correct secret key can open the message and use the encrypted public key of the sender to open the digital signature of the sender, etc. (Figure 2).

Figure 2 The use of SK and PKI to secure messages between users



In summary, the sender's encrypted message contains the original message, hashed value, and the corresponding public key. The recipient of the message will decrypt the encrypted message with the obtained public key from the sender. The original message is then passed through the hash function and a new summary string is computed. Hash functions operate on any variable-length message (King, 1997).

Hashing operates in a way that if the message content is changed then a new hash value is produced. The new computed digest is compared mathematically with the one attached to the sent message. If both are the same, then the recipient is assured that the message is not tampered with during transmission and that the integrity (hashing) and the authenticity (digital signature) of the message are preserved (Helms et al., 2002; King, 1997). However, this solution poses another sort of

complications. Although the hashed messages are not reusable but are subject to be resent. Repetition of a single message may be considered as multiple orders. One solution here suggests that the signature should include a sequence number as part of the hashing function, so that each transmission of the message generates a new hashed value, even if the message content remains the same (King, 1997). Then, the question is that how senders and receivers obtain public and private key combinations and who regulates such sensitive industry? Certified authorities, an in-house server or trusted third parties such as Entrust, Baltimore or Veresign, issue digital certificates to buyers and sellers over the Internet and consequently over mobile networks and devices (Levitt, 1998). The digital certificates include the public keys and identifying information of the person requesting the unique public/private keys. The wireless component is discussed next.

DIGITAL CERTIFICATE

Transactions conducted over wireless devices must be secured from one end to the other, therefore securing the mobile device itself, such as current WAP based phones or PDAs is not useful here because it does not secure the end to end connection passing through the whole wireless network infrastructure. Specifically, authentication and decryption of wireless networks and devices is less useful because they protect the radio interface only (Crowe, 2001).

Digital certificate, also called digital ID verifies that a public key belongs to a certain entity. Digital certificates are issued through a certificate authority (CA) as explained earlier. The CA verifies that a certain public key belongs to a specific company by binding the identifying information of the company or system to a public key with a digital signature. CAs schedule expiry date for certificates, responsible for revoking certificates and issuing Certificate Revocation Lists (CRLs) (Busta, 2002; Goldman, 2001; Hunt, 2001). The process of certificate request is handled through a Registration Authority (RA), which provides the interface between the user and the CA. The level of trust in the issued certificates relies on the strictness of the registration procedure. For example, if a RA requires an e-mail address and a name only from buyers/sellers then the trust of the issued certificates are expected to be very low.

Certificates are subject to change over time or/and to be compromised therefore users should be able to validate whether the certificate's data is current and even to revoke the certificate when necessary. Two methods are available to check the validity of certificates. Online validation allows users to ask the CA about a certificate's validity every time it is used and then the CA can simply state that the certificate is not valid. The other type of validation is called offline validation where the CA is asked to include a validity period in the certificate. With offline validation, Certificate Revocation List (CRL), which is a list of certificates that have been revoked before their scheduled expiration date, is the most common method to revoke certificates before the expiry date (Hunt, 2001).

Hunt (2001) states Private CA and public CA as the two major implementation models of CAs. Private CA is where providers such as RSA's kean 5.0, IBM's Secureway Trust Authority 3.1, sell a complete PKI system to an organization and it becomes the sole owner of the CA and responsible for issuing and managing certificates. Public CAs are certificates that are purchased from a public CA organization as and when required i.e., from Verisign.

PKI STANDARDS

The Internet Engineering task Force (IETF) is responsible for creating, promoting and disseminating standards for the Internet. With the evolution of the PKI and the need for security by Internet usets, IETF created working groups such as the IETF's PKIX group (Public Key Infrastructure X.509) to profile defined certificate formats for certain environments and uses and the ITU (the x.509 certificate) group to define certificate formats. The original ITU-X.509 certificate standard doesn't support many options for the contents of a certificate, while the x.509 profile defines the fields for the

Internet certificates exactly, in order to allow for better interoperability across different systems and applications. (Adams & Lloyd, 1997; Dankers, Garefalakis, Schaffelhofer, & wright, 2002; Hunt, 2001; Levitt, 1998)

The following table (Table. 1) presents the different profiles and protocols that have been standardised by PKIX. These are essential to manage and operate PKI. Table 1 states each specification, and provides an overview of each specification's activities (Adams, et al., 1997; Hunt, 2001).

Table 1 The different profiles and protocols of PKIX.

Specifications	Activities
PKIX-1 "Certificate and CRL Profile" (not a protocol specification) Profiles of the X.509v3 certificate standards. And the X.509v2 CRL standards for the internet.	-Define data structures associated with the X.509 certificates and CRLs -Doesn't specify a protocol -Represents certificates and CRLs as they would be found in a public repository such as an X.500 directory or other form of an enterprise directory service
PKIX-2 "Operational Protocols"	-Relying parties can obtain information such as certificates or certificate status
PKIX-3 "Certificate Management Protocols"	- Enable all aspects of certificate management - Include all the message elements which provide End Entity initialisation, certificate requests, key update, key recovery, and certificate revocation -Cross certification -CA key update -CRL publication -Messages that allow an end entity to request certain operating information from the CA
PKIX-4 "Certificate Policy and Certification Practice Statement Framework" (not a protocol specification)	-Presents a framework to assist the writing of certificate policy specifications and certification practice statements for CAs and for PKIs, listing many of the topics which may need to covered in such documents
PKIX-5 "Time Stamp Protocol"	-Define the protocol required to allow components to request a trusted time stamp, which is a necessary component in an environment that offers non-repudiation services.
PKIX-6 "Notary Protocols"	- Trusted and reliable third party notary services are essential to provide a mechanism to validate signatures and/or data relative to a given instance in time, also an integral component in order to facilitate true non-repudiation services.

It is interesting at this stage to introduce some of the commercial applications of some of the encryption tools and techniques highlighted earlier. Specifically, this research will discuss the Secure Electronic Transaction (SET) protocol, which utilises digital certificates.

SET

SET is a mechanism employed to secure electronic payment methods. SET utilises the public key certificate infrastructure to provide a secure environment for the transmission of sensitive data over the Internet, and assurances about the identity of the different parties involved in the process such

as customers, merchants and banks. SET relies on cryptography to ensure the confidentiality of the message, and allows for the use of digital signature to ensure the authenticity of users. SET creates digital signature through the use of distinct public/private key pair (Chaudhury & kuilboer, 2002; Minoli & Minoli, 1997).

Shoppers over the Internet may purchase goods in a variety of ways such as on-line catalogues, and electronic catalogues. SET protocol supports each of the previous shopping experiences as well as others, also the method of payment that the consumer over the Internet may choose varies thus SET supports a wide range of payment options and capabilities.

In the following, this research will attempt to link the above argument with mobile commerce security.

MOBILE COMMERCE

The security of wireless transactions involves securing the gateway server which sets on the top of the Internet as well as the handset. For example, 724 Solutions is an example of one vendor that introduced PKI gateway product, which is based on open standards, to allow organizations like banks to implement wireless PKI and digital signature capabilities that work with most PKI technologies and CAs on an increasing number of wireless devices.

PKI gateway solution allows carriers and other vendors to set policies, regarding which transactions that require digital signatures once integrated into the application framework solutions of that carrier. However, the PKI gateway solution doesn't solve the issue of device security (Goldman, 2001). Both SK and PKI techniques are utilised in securing mobile commerce. However, each has its own set of pros and cones. For example, the SK provides fast and efficient encrypt/decrypt way of authenticating messages, which makes them more suitable for devices of limited size and processing power, such as mobile phones and PDAs. Due to the availability of one key amongst the different users this jeopardises the whole security procedure as the number of people holding this unique key increases (for a group of n entities communicating with each other n(n-1)/2 keys are required) (Dankers, Garefalakis, Schaffelhofer & Wright, 2002). Above all, the secret keys need to be exchanged between users in a secure out-of-band mean in order to establish trust, which further complicates the situation and make it lengthier.

In the Global System for Mobile cellular radio (GSM) system the secret key is shared between the mobile subscriber and their home operator. This is achieved at the subscriber identity module (SIM) level which is owned by the mobile subscriber and administered in the database of the subscriber's home operator (Dankers et al. 2002). To overcome this problem, scientists came up with the PKI. The availability of one of the keys open to the public eliminates the need for the out-of-band procedure and hence, users does not have to establish trust prior to their exchanging of the keys. Thus, unlike SK, PKI does not require a lot of overheads but however needs an established CA infrastructure to distribute the public key authentically.

However, the mobile commerce (MC) environment is quite unique. The issue of pre-established security relations between providers and subscribers makes the possibility of integrating security with services based on SK. The providers of third generation (3G) mobile networks deliver smart cards with pre-installed symmetric keys, which are used to authenticate the mobile device (service subscription agreement) and the access network (via roaming agreements) (Dankers et al., 2002). Despite its superiority, PKI is not a favoured solution currently because issues such as non-repudiation and limited functionality of devices are not a strict requirement of wireless networks providers which make the SK more favourable than PKI, at least for the time being.

ADAPTING THE PKI TO SUITE MC REQUIREMENTS

According the above tradeoffs, PKI needs to be adapted to cope with the limitations of the mobile environment in order to provide end-to-end security. One of the solutions is the Wireless Transport

Layer Security (WTLS[1]) which is a PKI-enabled security protocol. It is designed for security communications and transactions over wireless networks (Dankers et al. 2002). It is used with the Wireless Application Protocol (WAP) to provide security on the transport layer between the WAP client in the mobile device and the WAP server in the WAP gateway (WAP1.2). The security services provided by the WTSL protocol are authentication, data confidentiality and data integrity. The WTLS protocol consists of a record layer protocol and a handshake protocol. To provide the authentication service and to generate the shared secret, the WTLS handshake protocol may use PKI for wireless environments (WPKI). WTSL provides functionality similar to the Internet Transport Layer Security systems (TLS) and Secure Socket Layer (SSL) but it has been optimised for use over narrow-band communication channel and incorporates datagram support in order to function at the transport layer. Also it does not support the Internet protocol (TCP/IP). Because WTLS and TLS are incompatible, content must be decrypted and re-encrypted as it passes through the WAP gateway. Recent developments on the WAP protocol (WAP2.0) eliminated the need for the WAP gateway. Hence, mobile WAP2.0 browsers support the Internet protocol (TCP/IP) directly and therefore, TSL can be supported by the handset and can run end-to-end from the handset to a Web server (not WAP gateway).

ISSUES

PKI is an ideal security solution yet a complex one, a combination of technologies, policies and processes. PKI's reputation has sharply declined during the past few years. PKI has not lived up to its promise. The following is a list of issues that have been the reason for the wide spread of disappointment with initiatives to deploy PKI (Corcoran, Sims, & Hillhouse, 1999; Hunt, 2001; Kolodzinski, 2002; Sundgren, 2002).

- Several PKI vendors such as RSA security, Entrust, and Verisign: no single vendor has been able to provide all PKI related technologies and services. Therefore, customers have to have a clear understanding of the major differences in approaches amongst the different vendors. Differences that can significantly affect how a PKI is managed and used.
- Scalability: depends on the relationship between CAs. Different methods may be followed in CAs to allocate trust and this problem increases as the certification path grows.
- Administrative burdens in deploying and in managing the PKI solution.
- Integration of the technology with actual application and systems in house.
- Maintenance problems.
- The process of setting up a PKI system.
- Secure key storage: public keys are widely distributed, often through a directory or database that can be shared by the public. These keys are typically a 1024-bit r 2048-bit string of binary digits with a unique property. Remembering these keys is impossible. Therefore, securing a storage place for them is essential.
- Cost: direct costs of the technology as the software and the hardware components such as the cryptographic modules are very expensive.

The above impediments are further aggravated by the following factors (Dankers et al. 2002):

- Mobile devices have limited processor capacity and memory storage. A high processing capacity is required for performing PKI operations and for the construction and validation of certificate chains. There is also a high demand for storage capacity to store the certificates and certificate revocation lists. One approach to overcoming the problem of limited device is to outsource some activities to a server.

[1] WTLS and Wireless Markup Language (WML) are optimised versions of HTML.

- Cross-certification is another issue where users from different PKIs are to be able to trust each other's certificates.
- Managing the complexity of a PKI in a limited bandwidth represents another challenge. However, the radio link will probably not be a problem in the future (i.e. 4G).
- Interoperability issues represent a challenge here. The existence of different certificate formats, which are tailored to the environments in which they are used. This could be resolved by specifying certificates with as few parameters as possible. Even if the same format is used, interoperability problems may arise due to the certificate extensions that are defined. For example, if an entity receives a certificate with an extension marked as critical and does not understand the extension, the certificate is rejected, and this complicates the security functions. Restricting the use of extensions and the use of the criticality flag could enhance this. There are other incompatibilities using the certificates.
- Organisational issues do possess a challenge here represented by the several means available to generate and store the PKI key pair (e.g. mobile device itself, manufacturer of the device, SIM, smart card).

Sundgren (2002) argues that in spite of all the above issues and the many PKI deployment failures, there are some success stories and several trends that support a brighter future for PKI. The following are some of the factors behind the optimistic predictions (Sundgren, 2002):

- Improving technology: technologies such as smart cards and readers are coming down in price which address issues like secure storage places of the private keys and make this approach to storing private keys more feasible.
- Easier application integration: PKI enabled applications continue to increase. PKI vendors offering more and more out of the box solutions that yield immediate benefits.
- Increasing demand: PKI capabilities can be provided by other technologies, but the fact that PKI offers a coherent framework for achieving multiple functions in a coordinated way leads to increase the demand for PKI as the demand for security related capabilities increases.
- Web services: security requirements for web services could lead to interest in PKI. SSL provides a basic security capability, which is not enough for web services as they evolve.

CONCLUSION

Thus far, the use of PKI in wireless devices and networks is limited currently because, there will always exist a need for a pre-established relationships (trust) between the service provider and the mobile subscriber. SK cryptography is an ideal solution for the time being. The limited hardware functionality of mobile devices and bandwidth capacity represents another problem for PKI deployment in MC. However, it is believed that the future holds great promise for PKI as the technology enhances. The potential advantage of PKI in the future is that mobile devices can access wireless networks based on spontaneous networking protocol (TCP/IP) where PKI and trusted third parties could come into play directly without the hassle of going through e.g. WAP gateways.

REFERENCES

Adams, C., & Lloyd, S. (1997, 29-31 Oct). Profiles and protocols for the internet public-key infrastructure. Paper presented at the Proceedings of the sixth IEEE Computer Society Workshop on Future Trends.

Busta, B. (2002). Encryption in theory and practice. The CPA journal, 72(11), 42-48.

Chaudhury, A., & kuilboer, J.-P. (2002). e-business and e-Commerce infrastructure. New York: McGraw-Hill.

Corcoran, D., Sims, D., & Hillhouse, B. (1999). Smart Cards and Biometrics: your key to PKI. Linux journal, 1999(59es), 1-7.

Crowe, D. (2001). Simplified security. Wireless review, 18(20), 37-38.

Dankers, J., Garefalakis, T., Schaffelhofer, R., & wright, T. (2002). Public key infrastructure in mobile systems.

Electronics and communication engineering Journal, 180-190.
Goldman, C. (2001). Banking on security. Wireless Review, 18(7), 22-24.
Helms, G. L., Bushong, J. G., & Nelms, L. (2002). Security in internet e-commerce. CPA Journal, 61(3), 12-15.
Hunt, R. (2001). PKI and digital certification infrastructure. Paper presented at the Networks, Ninth IEEE International Conference on, 2001.
King, C. M. (1997). Public key infrastructure: End-to-end security. Business communications review, 27(11), 50-54.
Kolodzinski, O. (2002). PKI: commentary and observations. The CPA, 72(11), 10-11.
Levitt, J. (1998). The keys to security. Informationweek (698), 51-60.
Minoli, D., & Minoli, E. (1997). Web commerce technology handbook. New York: McGraw-Hill.
Sundgren, J. (2002). Market overview: Public key infrastructure: Giga information group, Inc.