Security Measures in Mobile Commerce: Problems and Solutions

Sanwar Ali
Waleed Farag
Mohammad A. Rob

Follow this and additional works at: https://aisel.aisnet.org/iceb2004

This material is brought to you by the International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICEB 2004 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Security Measures in Mobile Commerce: Problems and Solutions

Sanwar Ali\textsuperscript{1}, Waleed Farag\textsuperscript{2}, Mohammad A. Rob\textsuperscript{3}

\textsuperscript{1}Department of Computer Science, Indiana University of Pennsylvania, Indiana, PA 15705, USA sanwar@iup.edu, Fax: 724-357-2724
\textsuperscript{2}Department of Computer and Systems Engineering, Zagazig University, Zagazig, Egypt farag@cs.odu.edu, Fax: +20-55-2304987
\textsuperscript{3}Management Information Systems, University of Houston- Clear Lake, Houston, TX 77058, USA rob@cl.uh.edu, Fax: 281-283-3951

ABSTRACT

Due to the advent of the Internet, electronic business transactions have exploded around the globe. Along with the Internet, wireless technology has exponentially developed as well. Today, new technologies that allow mobile (cellular) phones and other handheld devices to access the Internet have made wireless business transactions possible. This phenomenon is known as mobile commerce or M-Commerce. It has been predicted that the number of mobile phones connected to the mobile Internet will exceed the number of Internet-connected PCs before 2007. The mobile phone will therefore become the most prevalent device for accessing the Internet. Several industry analysts predict that M-commerce will constitute a multibillion dollar business by 2005. However, M-Commerce brings new challenges in providing information security as information travels through multiple networks often across wireless links. What must be done to secure financial transactions via mobile commerce? Generally speaking, M-Commerce creates more security concerns than traditional E-Commerce. In this paper, security measures in M-Commerce, wireless security, and the application of cryptography for key generation, authentication, digital signature and digital certificate are discussed.

Key words: M-Commerce, wireless security, elliptic curve cryptography, WLAN security, Bluetooth

1. INTRODUCTION

Mobile commerce, commonly known as M-Commerce, ‘is concerned with the use, application and integration of wireless telecommunication technologies and wireless devices’, such as Internet-enabled mobile phones, personal digital assistants (PDA), palmtops, laptops, and pagers, commonly know as handheld wireless devices (HWD). The discipline of M-Commerce includes reference to the infrastructures and electronic technologies necessary for wireless data and information transfer in the form of text, graphics, voice, and video. M-Commerce is a subset of electronic commerce where the Internet-enabled HWDs and wireless networking environment are necessary to provide ‘location independent connectivity’. It is predicted that M-Commerce services would be the next biggest growth area in the telecommunications market, represent-ing the fusion of two of the current consumer technologies: wireless communications and E-Commerce [7].

Mobile communications is now considered a relatively mature technology with the transition from second to third generation (3G) wireless technology and with its high consumer acceptance. 3G wireless technology provides a variety of services and capabilities in addition to voice communication, such as multimedia data transfer, video streaming, video telephony, and full Internet access. 3G mobile phones provide high-speed data transfer and are designed to support large numbers of users more efficiently allowing for future expansion in user capacity. This technology was first introduced in Japan in 2001 and then spread in Europe and USA in 2002. In Japan mobile commerce has been a huge success, accounting for over $400 million in revenues each year [8]. In 2003, there were about 800 million Web-enabled cellular phones in use worldwide. By 2004, the worldwide number of Web-enabled mobile phones is projected to rise to one billion. The M-Commerce market may rise to $200 billion by 2004 (Gartner Group). It has been predicted that the number of mobile phones connected to the mobile Internet will exceed the number of Internet-connected PCs before 2007. The mobile phone will therefore become the most pre-valent device for accessing the Internet and M-Commerce. Several industry analysts predict that mobile commerce will constitute a multibillion dollar business by 2005. However, once again, the question of security arises. What must be done to secure financial transactions via mobile commerce? Secure mobile commerce will be a key area within the expanding functionality of mobile phones. Generally speaking, M-Commerce creates more security concerns than traditional E-Commerce.

2. APPLICATIONS OF M-COMMERCE

There is a wide variety of applications of M-Commerce. For example, Smart cards with an embedded integrated microchip can be used as prepaid phone cards, ATM cards, or public transportation cards. They can be biometrically enhanced to include voice recognition, iris and face scans, and finger print authentication. In France, some 35 million Smart cards are in circulation and every year they process over three billion transactions [4]. Wireless banking refers to purchasing over Internet-enabled HWDs like wireless application protocol (WAP)
phones or PDAs. Interactive TV is enhanced TV where additional content is added to an existing broadcast format that the viewer can query, request or even interact live with the program. It has reserved channels and bandwidth for data applications such as weather, news, games, or commerce. It also offers services like Video on Demand (VOD) or Personal Video Recording (PVR). Companies are providing facilities to track stocks on HWDs. Aspiro, a Swedish company, allow its customers to check stock prices or look at their portfolios and even trade using WAP phones or PDAs. Accessing information using WAP mobile phones and PDAs is significantly becoming popular for business-to-business (B2B) and business-to-consumer (B2C) applications. Mobile and wireless technologies have been an integral part of defense and military ever since these technologies were available. In light of growing concern with cyberterrorism, security has undoubtedly become the single most important issue because many U.S. intelligence services use mobile technologies and wireless networks for communication and commerce.

3. M-COMMERCE SECURITY CONCERNS

M-Commerce is bringing together two technologies, wireless communication and traditional E-Commerce, with a history of security problems. Coupled with the convergence of voice and data communications, interconnection with external data networks and issues surrounding the transactions themselves, the potential risks are very high [7]. There are three basic security components in M-Commerce: (a) Transaction: protecting the transaction parties and their data by providing an acceptable level of security, (b) Information: protecting valuable and sensitive information about customers, and (c) Infrastructure: protecting the network infrastructure from attack.

Mobile 3D is Visa International’s new global secure specification that is expected to ensure the security of Internet payments made over mobile devices [14]. It was developed in conjunction with some 15 industries, including Ericsson, Motorola, and Oracle Mobile, and was launched in September 2001. It extends payment authentication initiatives into M-Commerce, enabling the Visa card issuers to validate the identity of their cardholders in real time, ensures that payment data sent over open networks is not compromised and allows cardholders to actively protect their Visa accounts from unauthorized use, and supports global interoperability, enabling consumers to have a consistent and seamless experience regardless of the method or device being used to access the Internet. A number of Visa M-Commerce programs are currently underway worldwide to test the validity of M-Commerce payment solutions and raise consumers awareness. In Asia, Visa has partnered with Hutchison Telecommunications and Dao Heng Bank to develop a mobile payment service using Mobile 3D Secure. In Europe, Visa has made a strategic alliance with Omniteq Vodafone, while in the US, Visa and Sprint are working together to help facilitate secure mobile payments.

A satisfactory level of security is required for the successful deployment of HWDs. To a certain extent, it seems reasonable to utilize solutions used for the wired environment in the case of the wireless environment. However, this approach is not always feasible because of the differences between the wired and wireless environments. For example, because of the hardware limitations of the HWDs, no large routing tables can be maintained on these devices, thus increasing the risk of a denial-of-service attack. Furthermore, wireless communications make physical eavesdropping almost undetectable [10]. Similar to wired communication, wireless communication also needs three basic security requirements: (i) confidentiality- information is disclosed only to legitimate entities or processes, (ii) integrity- unauthorized modification of information is prevented, and (iii) availability- authorized entities can access a service provided they have approved privileges. The mechanisms suitable to the HWD are discussed here.

M-Commerce needs several layers of security: (i) device security, (ii) language security, (iii) wireless security, and (iv) cryptographic security.

3.1 Device Security

Within the design of mobile devices, there are a number of high quality security features. The most important of these are: (i) a built-in password mechanism which will lock after several mistyped attempts and (ii) an industry approved, tamper-proof smart card, known as Subscriber Identification Module (SIM) card.

The SIM card and the mobile device are always stored together and the device is an every day utility object that is easily lost or stolen. Time-out and key-locks are often not used on phones. This means that as long as the phone remains turned on the strong password system will be by-passed. All WAP data, in some popular handsets, is stored in the phone’s memory, not in the SIM; this will include login and password information. These features certainly diminish the security of the mobile phone. The SIM card used in mobile devices are de facto microprocessor and they can be used to facilitate mobile commerce. Gemplus SIM cards features a digital signature and public key encryption [11] and the technology is embedded in the card. In May 1999, Motorola, jointly with Identix, a biometrics company, developed fingerprint scanning devices, called the DFR 300 that is 4.5 millimeters thick [2]. This scanning device can now be incor-porated into the HWDs.

3.2 Language Security

If special purpose M-Commerce software, such as a stock trading application, is to be deployed on mobile
devices, then Java is the recommended language to be used as the deployment language on the HWDs. By using Java, the amount of software that needs to be changed in order to adopt the application to various mobile platforms is minimized. Feasible Java execution environments are available for PDAs, Smart phones, Communicators (such as Symbian), laptops, and other platforms. Maffeis [6] also recommended using server side Java technology, such as the Java-2 Enterprise Edition (J2EE) platform, in the data center. This allows for shorter time-to-market and avoids vendor lock-in.

3.3 Wireless Security

3.3.1 WAP Security

WAP (Wireless Applications Protocol) is ‘an open, global specification that empowers mobile users with wireless devices to easily access and interact with information and services instantly.’ WAP is currently the only publicly available solution for wireless communication and enables M-Commerce where Internet data moves to and from wireless devices. WAP-enabled phones can access interactive services such as information, location-based services, corporate information and interactive entertainment. WAP is targeted at various types of HWD and Bluetooth enabled mobile phones.

WAP 1.x security uses the Wireless Transport Layer Security (WTLS) protocol. This protocol is the WAP equivalent of Secure Socket Layer (SSL) and it provides authentication, encryption and integrity services. WTLS has three levels, all have privacy and integrity: (i) Class-I has no authentication (anonymous), (ii) Class-II has server authentication only, and (iii) Class-III has both client and server authentication. WTLS supports some familiar algorithms like Diffie-Hellman, RC5, SHA-1, and IDEA. It also supports some trusted methods like DES and 3DES, but it does not support Blowfish and PGP [9].

Since Web- and WAP-based protocols are not directly interoperable, a component known as the WAP gateway is needed in order to translate Web-based protocols to and from WAP-based protocols. The WAP gateway is a software that runs on the computer of the Mobile Service Provider (MSP). Thus sensitive information is translated into original unencrypted form at the WAP gateway [5]. This problem is known as WAP gap.

Public key cryptography (PKC) is used to exchange a symmetric or private key using certificate and then all transmission is encrypted. A short key size of 40 bits is used because of power limitation. A tamper-proof component, know as WIM (Wireless Identity Module) is designed as part of the WAP architecture to store private data, such as key pairs, certificates, and PIN numbers within the mobile device. In practice, a WIM is implemented using a smart card. Wireless Markup Language (WML) is used in WAP 1.x technology. Figure 1 shows WAP gap model.

WAP 2.0 security uses TLS (Transportation Layer Security) instead of WTLS due to requiring end-to-end security with all IP based technology in order to overcome the WAP gateway security breaches. It is a Public Key Infrastructure (PKI) enabling protocol that provides the services such as authentication by using digital signatures and public key certificates, confidentiality by encrypting data, etc. This protocol uses RSA, RC4, 3DES, and SHA-1 algorithms for encryption. Wireless PKI (WPKI) is released for the first time. Figure 2 shows the WAP proxy model.

PKI systems and WTLS are at the heart of today’s mobile security technology. In a WAP environment WTLS must be translated at the WAP gateway into SSL, the Internet standard. A PKI is a set of policies, processors, software, hardware, and technologies that use PKC and certificate management to secure communication [18]. PKI’s trusted services enable the secure transfer of information and supports a wide variety of M-Commerce applications. PKI must ensure the following: (i) confidentiality, achieved by cryptography, (ii) authentication, achieved by digital certificates, (iii) integrity, achieved by digital signatures, and (iv) nonrepudiation, achieved by digital signatures and certificates.

PKI consists of the following components: (i) Certificate Authority (CA)- responsible for issuing and revoking certificates, (ii) Registration Authority (RA)- binding
between public key and the identities of their holders, (iii) Certificate Holders- people, machine or software agents that have been issued with certificates and can use them to sign digital documents, (iv) Verification Authority (VA, Clients)- validate digital signatures and their certificates from a known public key of a trusted CA, and (v) Repositories- stores that make available certificates.

WPKI is an optimized extension of traditional PKI for the wireless environment. ‘WPKI encompasses the necessary cryptographic technology and a set of security management standards that are widely recognized and accepted for meeting the security needs of M-Commerce’ [18]. WPKI applications have to work in an environment with less powerful CPUs, less memory, restricted power consumption, and smaller displays. WPKI solutions are likely to employ “network agents” that take care of some of these tasks. The HWD must at least be able to perform a digital signature function to permit the establishment of a WPKI. Network agents can perform all other WPKI-related tasks such as validation, archiving or certificate delivery. Figure 3 shows a schematic diagram of WPKI.

The private keys are stored in WIM or SWIM. Two main components of WPKI are the PKC and the key certification management. In order to perform globally, worldwide PKI legislation is required.

A number of companies, including Entrust, Certicom, RSA Security, VeriSign, and Baltimore, have announced solutions enabling the use of PKI software in a wireless environment. The biggest challenge in implementing a PKI solution for wireless deployment lies in the devices them-selves. With limited bandwidth and low power, a small screen and no keyboard, the average mobile device presents a number of unique problems. Certicom developed Elliptical Curve Cryptography (ECC) that reduced the key size from RSA’s 1024 bits to as few as 56 bits and made handling certificates a lot easier for low-bandwidth and low-power devices [2]. This technique can process a digital signature in a second compared to the RSA technique, which takes almost 15 seconds to process a digital signature on a Palm device. Both Nokia and Ericsson have been supporting server certificates since the middle of 2000.

3.4 Wireless LAN (WLAN) Security

3.4.1 IEEE 802.11b

The WLAN standard IEEE 802.11b provides a mechanism for authentication and encryption. It provides a maximum of 11 Mbps wireless Ethernet connections using the band at 2.4 GHz. 802.11b security features consists of security framework called Wired Equivalent Privacy (WEP). WEP is based on RC4, a symmetric stream cipher. It has a pseudo-random number generator, whose output is XORed to the data. WEP can use 40 or 128 bits key size. However, using a 128 bits key size, 802.11b throughput drops much due to heavy calculations. In August 2001, RC4 was announced to be broken and can be cracked in less than half an hour. Consequently, WEP can be broken. WEP with 40 bits key size can be broken in real time.

3.4.2 Bluetooth

Bluetooth technology, developed by Ericsson in 1998, is used to connect different HWDs and provides a method for authenticating devices. Device authentication is provided using a shared secret between the two devices. The common shared secret is called a link key, generated from PIN. This link key is established in a special communication session called pairing. All paired devices share a common link key. There are two types of link keys: (i) unit keys and (ii) combination keys [17]. The link key is a 128-bit random number.

A device using a unit key uses the same secret for all of its connections. Unit keys are appropriate for devices with limited memory or a limited user interface. During the pairing procedure the unit key is encrypted and transferred to the other unit. Only one of the two paired units is allowed to use a unit key. Combination keys are link keys that are unique to a particular pair of devices and they are only used to protect the communication between these two devices. Clearly a device that uses a unit key is not as secure as a device that uses a combination key. Since a unit key is common to all devices with which the device has been paired, all such devices have knowledge of the unit key. Consequently they are able to eavesdrop on any traffic based on this key.

In every Bluetooth device, there are four entities used for maintaining the security at the link level: (i) the Bluetooth device has an IEEE defined 48-bit unique address, (ii) a private authentication key which is a 128-bit random number, (iii) a 8-128 bit long private encryption key, and (iv) a random number, which is frequently a changing 128-bit number that is made by the Bluetooth device itself [12]. The security algorithms of Bluetooth are considered strong. Bluetooth standard does not use the RC4 cipher; rather it uses the E1, a modified block cipher SAFER+. No practical direct attack has been reported.
4. CRYPTOGRAPHY

The two types of cryptography currently available are (a) symmetric or secret-key and (b) asymmetric or public key cryptography (PKC). In secret-key cryptography, two devices must share their secret key in order to communicate securely. Thus two concerns arise: how to exchange the secret key securely; and if $n$ HWDs must communicate with each other, a total number of $O(n^2)$ secret keys must be exchanged. The management of such a number of secret keys should consider the scalability issues.

In PKC, both the above problems are solved since two parties in the communication do not require exchanging any secret keys. According to this consideration, PKC seems to be the ideal candidate to enforce confidentiality. Cryptographic techniques already exist to protect data transmission over the Internet. The de facto cryptographic algorithm for digital signatures and for encryption of secret keys is the RSA PKC. For electronic commerce, it is a highly secure technique. Although RSA is highly secure and widely used, there are some potential problems with its use in M-Commerce: (i) in RSA the key size is large which requires a significant amount of memory for storage, (ii) decryption time increases rapidly as the key size increases, and (iii) key generation is complex and time consuming. Memory constrained devices cannot easily generate RSA keys and therefore may need to have keys generated by another system. Moreover, the computations needed to encrypt and decrypt messages using RSA is overwhelming compared to secret key cryptography.

4.1 Elliptic Curve Cryptography (ECC)

What is really needed is a public key algorithm that achieves a high level of security using short keys. Algorithms based on mathematical objects known as elliptic curves offer interesting possibilities [1]. Elliptic curve discrete logarithm problems (ECDLP) is defined as “give a base point $P$ and the $kP$ lying on the curve, find the value of $k$”. From cryptographic point of view, a new crypto-graphic system needs to be defined based on elliptic curves. Any standard system that relies on the discrete logarithm problem has a direct analogy based on the ECDLP. For example, Elliptic Curve Digital Signature Algorithm (ECDSA) has already been standardized. Diffie-Hellman key exchange can be easily implemented in an elliptic curve system.

Certiocom promotes the use of ECC, as well as its mobile Virtual Private Network (VPN) client, and a client, which bring sophisticated security functionality to resource-constrained devices. WPKI ECC is recognized as the most optimized and therefore the best suited for supporting security in the wireless environment. The key for digital signature in ECC is 163 bits as compared to 1024 bits in other signature schemes. However, good elliptic curves must be carefully chosen, otherwise it might be prone to various attacks. ECC is small, efficient and requires low power. Table 1 shows key sizes for ECC and RSA.

<table>
<thead>
<tr>
<th>ECC Key Size (bits)</th>
<th>Traditional RSA Key Size (bits)</th>
<th>Key Size Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>109</td>
<td>512</td>
<td>1:5</td>
</tr>
<tr>
<td>131</td>
<td>768</td>
<td>1:6</td>
</tr>
<tr>
<td>163</td>
<td>1,024</td>
<td>1:6</td>
</tr>
<tr>
<td>283</td>
<td>3,072</td>
<td>1:11</td>
</tr>
<tr>
<td>409</td>
<td>7,680</td>
<td>1:19</td>
</tr>
<tr>
<td>571</td>
<td>15,360</td>
<td>1:27</td>
</tr>
</tbody>
</table>

Recently, ECC has been deployed on Smartcards without coprocessors [16]. Weimerskirch et al [15] implemented ECC on a Palm OS device. Their study showed that the normal transaction, such as a key exchange or signature verification, can be done in less than 2.4 seconds while signature generation can be done in less than 0.9 seconds. Table 2 shows the three major industry standard PKC systems that can be considered secure, efficient, and commercially available [13].

<table>
<thead>
<tr>
<th>PKC</th>
<th>Mathematical Problem</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integer factorization</td>
<td>Given a number $n$, find its prime factors</td>
<td>RSA, Rabin-Williams</td>
</tr>
<tr>
<td>Discrete logarithm</td>
<td>Given a prime $n$, and numbers $g$ and $h$, find $x$ such that $h = g^x \mod n$</td>
<td>ElGamal, Diffie-Hellman, DSA</td>
</tr>
<tr>
<td>EC discrete logarithm</td>
<td>Given an elliptic curve $E$ and points $P$ and $Q$ on $E$, find $x$ such that $Q = xP$</td>
<td>EC-Diffie-Hellman, ECDSA</td>
</tr>
</tbody>
</table>

4.2 Problems with Elliptic Curve System

The true difficulty of the ECDLP is not yet fully understood [1]. Recent research has shown that some elliptic curves that were believed suitable for ECC are, in fact, not appropriate. For example, if the order of the base point $P$ is equal to the prime number $p$ then it turns out that the ECDLP can be solved efficiently. Such curves are known as anomalous curves. For a given curve and base point, it is trivial to generate public and private keys (the private key is simply a random integer $k$ and the public key is the point $kP$ on the curve). However, it is an extremely difficult problem to generate a suitable curve and base point in the first place.

The main problem is how to count the number of points on the curve. Having done this, it is then necessary to select a suitable base point $P$, which must have a large order to ensure the difficulty of the ECDLP. But the order of $P$ must divide the number of points on the curve.
Having found the number of points on the curve, it is quite likely that a suitable base point cannot be found. Users may use random curves or special curve generating software, such as the “Elliptic Curve Generation Bureau” created by Zaxus.

In April 2000, the French National Institute for Research in Computer Science and Control announced that the 109-bit ECC key was cracked in a four-month brute-force effort using 9,500 computers by 1300 volunteers from 40 countries [3]. The amount of work required was about 50 times that required to solve a 512-bit RSA cryptosystem. The NIST has endorsed 163-bit ECC and WAP standards to be used in HWDs.

4.3 Digital Certificates and Digital Signature

In a PKC, a message is encrypted using a public key and is decrypted using a secret key. However, there is no inherent way of knowing the person who has the corresponding secret key. This is where the idea of certificates arises. Certificates confirm that the public key given in the certificate belongs to a private key held by the legitimate person, not by an imposter. To trust a certificate means to trust the party who issued the certificate, not the person for whom the certificate is issued. To protect a certificate from being modified one uses digital signatures. The message can only be created in an inherent way of knowing the person who has the corresponding secret key. This provides authorization and non-repudiation. That is the basis for digital signature.

How to protect private keys? The private key is stored in a Smart card, where all crypto operations with it are performed. The Smart card access gets restricted by the use of a PIN. The place where a user’s private credentials are stored is called Personal Security Environment (PSE).

5. CONCLUSION

M-Commerce security is a very crucial issue that needs further research to introduce efficient and effective solutions. In this article, various security concerns were expounded. ECC certainly appears to provide a viable alternative to RSA. There are potential advantages, especially when used in devices with limited processing capability and memory. Typical applications include M-Commerce using handheld wireless devices. There are, however, some problems and issues that are inhibiting the widespread adoption of ECC. These include (i) the real security of such systems is still not well understood, (ii) difficulty of generating suitable curves, and (iii) relatively slow signature verification. Time will tell its future.

REFERENCES