Enhanced SIM (ESIM): a proposal for mobile security

Giuseppe "Gippa" Paternó
Centre for Telecommunications Value-Chain Research
Trinity College Dublin
Dublin 2, Ireland
paternog@cs.tcd.ie

September 2009

***Abstract*—The Enhanced SIM wish to address primarily the need of an user to be uniquely identified within corporate and Internet resources. By leveraging a public key infrastructure of a mobile operator, web sites and corporation can federate with the operator to authenticate the end user by reusing the technology that is already available. As a side effect, a number of VAS applications -such as bank services- can take advantage of this feature to secure financial transactions over mobile devices.**

I. Introduction

The increasing number of Internet web sites and corporate applications increased also the username and password combinations a user should keep track. Some corporate access and financial institutes even uses OTP devices: it happens that a user can have multiple OTP devices, making uncomfortable to bring all along. Although some mechanism of single sign-on and of federation has been adopted, yet the number of OTP devices and username/password combinations are too high.

The physical identity of a user doesn't change whatever he/she uses a given services: the idea is then to leverage something that a user always brings with him/her. Almost everyone has at least one mobile phone nowadays, making the phone an ideal device to identify an user. The Enhanced SIM wish to address the need of identifying uniquely the user to every web site and resource by leveraging the X.509 certificates and their corresponding private certificates.

As a side effect, a number of other services can be accomplished by integrating a security infrastructure into the phone, such as financial services and other value added services (VAS). In fact, the result of the continuous evolution of the mobile telephone networks and the users' needs, the mobile phone is becoming a terminal to access both personal and corporate data, such as bank/financial services.

Financial and bank services access through the mobile are becoming more popular among consumers and small business: thanks to the introduction of the Mobile Virtual Network Operators (MVNO), many banks creates their own MVNO giving access to their bank services through applications stored on the SIM itself and created with the SIM Application Toolkit. As applications made with the SIM toolkit are basically exchanging Short Messages (SMS) in plain-text, these services are most likely to be a good target from malicious attackers.

The Internet is facing an increasing number of threats both in the server and the client sides. Attacks are getting more and more sophisticated, involving networks, operating systems and applications. As the users are using their mobile devices as a personal and corporate terminal, we expect that such threats will move into the mobile space sooner or later.

Mobile telephone networks are mostly based on GSM (2G) or UMTS (3G) networks and are considered quite trusted, however we expect that next generation mobile networks will be mostly IP-based such as leveraging WiMAX or similar layer 2 network medium, with the consequences that IP-based networks can have in the security issues.

II. The Enhanced SIM (ESIM)

Current mobile telephone networks use a Subscriber Identity Module (SIM) or an Universal SIM (USIM) that is a special smartcard containing its unique serial number, international unique number of the mobile user (IMSI), security authentication and ciphering information, temporary information related to the local network (also temporary local id that has been issued to the user), list of the services the user has access to and two passwords (regular PIN and unblocking PUK).

The SIM smartcard is composed of directories and files that are described in detail in the ETSI GSM Technical Specifications 11.11[1] and 11.14[2]. The file system may be comprised of the following basic forms: a master file (MF)[3], a directory file (DF)[3] and an elementary file (EF)[3]. The subscriber and operator information are hold within this structure.

This paper proposes embedding a cryptographic engine in a SIM to hold X.509 certificates and the corresponding private cryptographic keys, to be used for authenticating to external or remote applications. When shipped, the SIM will hold a key-pair generated at the personalisation stage by the mobile operator and signed with the operator Certification Authority. This implementation of a SIM is referred from now on as Enhanced SIM (or ESIM).

III. Smartcard specifications

Recent studies of the author [4] state that the smartcard market for authentication is unclear at the moment. While the

standard for the reader access is focused on the well-accepted CCID specification, middleware access is not standardized yet: each manufacturer need its own middleware with a proprietary PKCS11 or CSP library to interface with a cryptographic engine.

The GSM market needs a well established standard to interface with the SIM, as it represent one of the core part of the GSM network itself. Most of the smartcard manufacturer are converging over the JavaCard specifications[5], and its specification should represent the minimum standard over the Enhanced SIM implementation.

However, the Java Card technology provides an architecture for open application development for smart cards, using the Java programming language. Basically, the smartcard contains a Java Virtual Machine (JVM) capable of running java code on the card microprocessor; this Java code is referred to as an applet. A Java applet therefore should provide cryptographic function to the middleware in the mobile device.

In an ESIM, the identified applet is the Musclecard applet as defined in the *MUSCLE Cryptographic Card Edge Definition for Java Enabled Smartcards specifications*[6][7][8].

The card manufacturer, upon request of the operator, can optionally provide dual contact and contactless cards so that leveraging the JavaCard specifications will open the card for other possible application scenarios.

IV. Certificate lifecycle

A. Certification Authority

Each operator *must* have a Certification Authority (CA), who is responsible of signing the certificate on-board the ESIM and to manage certificates lifecycle. It is strongly suggested that the CA follows the common criteria specifications.

Each operator has to publish the following CA information:

- CA root certificate URL
- The URL of the Certificate Revocation List (CRL) for off-line verification
- Online Certificate Status Protocol (OCSP) responder (optional) for on-line verification

The place in which those information are stored must be made publically available to other operators. It is strongly suggested that the ESIM preload the CA public certificate of the operators with which it has relationship. The mobile device has to retrieve CRL URL and OCSP responder information from the CA certificate. Furthermore, the public key for lawful interception certificate has to be clearly specified, as discussed later in section VIII.

The mobile device should update regularly the associated CRLs over a well-defined Access Point Name (APN). The mobile operator can optionally send its own CRL via OTA Provisioning. The CRL is a key part of ensuring that the peer certificate has not been revoked. Whenever a data connection is available, the mobile device should try attempting the OCSP verification (when available by the operator) first.

B. ESIM certificate

When a SIM is created and personalized for the operator at the factory, it will be pre-loaded with an X.509 key-pair.

A key generation command will be issued to the on-board java applet, the public key submitted to the operator's CA and the signed public certificate loaded on the card.

The certificate subject will have the *common name set to the card IMSI value*, other values are optional but must reflect the Certification Authority values. For example, organization (O) must reflect the mobile operator name and the country code (C) set to the one specified by the IMSI value.

Certificate renewals will be fulfilled using the SCEP protocol.

C. Provisioning

The certificate lifecycle can be easily integrated into an operator provisioning system. When the SIM is issued, the key-pair is generated in the factory and then signed with the operator CA. If the SIM is stolen or replaced, the certificate will be revoked from the operator.

V. APIs access

Each device manufacturer must publish APIs to access the keys infrastructure, either to internal applications as well as third parties, if possible leveraging and integrating with the existing on-board certificate management. For example, Windows Mobile devices can leverage the CertificateStore APIs[9] or iPhone the keychain service[10].

By leveraging these APIs, third party application that already uses these APIs can integrate the on-board certificate. For example, on-board Web Browsers can optionally permit certificate login to a server whenever required, a VPN software can use the certificate for login purposes and device storage can be encrypted with the user public key.

VI. Cryptographic engine access

Mobile devices should expose crypto services to standard computers. Through this feature, the phone device and the SIM can be used to access corporate resources and public sites.

The manufacturer should expose the ESIM applet by emulating a standard smartcard reader. When this feature is available by the manufacturer, the mobile device must have a cable that is able to connect to a standard USB port interface in the computer and emulate a smartcard reader following the USB CCID specifications. The advantage of USB CCID specification is that almost well accepted across multiple operating systems such as Linux, Microsoft Windows, Sun Solaris and Apple Mac OS X.

VII. Applications

The ESIM together with API access within the device can address a large number of applications that can take advantage of the security infrastructure provided.

This chapter, however, don't wish to address the specifications of the application themselves, rather wants to give examples on how this security infrastructure can be deployed in common scenarios.

A. Short Messaging Service

Although operators' core telephone network is considered somehow trusted, it is possible to spoof the source of the short message. Most banks, especially those who also owns a Mobile Virtual Network Operator (MVNO), uses the SIM Application Tookit to query bank services: the toolkit leverage SIM to exchange SMS with the services to perform the requested actions.

SMS are also used as an authentication mechanism that send a One Time Password (OTP) to the end-user. Spoofing SMS can lead to data leaking in these areas or even misuse the device, as the attack described in [11].

Using certificates to sign SMS can ensure that the message is sent by a valid entity in the telephone network and is not spoofed, ensuring that the source is verified. To interactive messages, the mobile device should display to the end-user that the message is signed and verified.

Whenever a signature is applied to the SMS, all services that relies on the Short Message Service, including applications based on the SIM Toolkit, must verify that the signature is valid. SIM Application Toolkits can also allow to encrypt data using a specified public key that can be included in the application.

Note that signing SMS means that the message will not fit a single short message that is limited to 160 characters. SMS reassembly is a critical part of the process: a single message dropped means that the whole content is lost. This has to be taken into consideration when designing applications with the SIM Toolkit and on the receiver side.

B. Remote Device Wiping

Devices and laptops are sometimes target of criminal organizations to gather precious information about a company, such as copying confidential information. Even an unintentional loss can represent a damage to the person that holds private data on board.

By leveraging the certificate system, an operator can send a special short message for wiping the device remotely. The issuing operator will send a special signed SMS that contains instruction for the device for an immediate wipe. As the message is received at the mobile terminal, the device must immediately wipe its memory, without prompting the user and even if it's in standby.

C. IP-based telephony

The convergence to IP-based telephony, widely adopted now in landline communication, is expected to have the same spread adoption when next generation mobile networks such as WiMAX will be available.

The certificates can be used to secure IP-based phone communication, as the current weakness of the protocol is that an attacker can listen to the conversation with easy to find tools such as *Cain & Abel*[12]. For example, ESIM can be used to protect the SIP protocol with TLS[13] or generate a random AES key in the Secure Real-time Transport Protocol (or SRTP)[14]. The SRTP defines a profile of RTP (Real-time Transport Protocol) that addresses encryption, message authentication and integrity, and replay protection to the RTP data.

D. IEEE 802.1x

IEEE 802.1X is an IEEE Standard for port-based Network Access Control and it's part of the IEEE 802.1 group of networking protocols. It provides an authentication mechanism to devices wishing to attach to a LAN, either establishing a point-to-point connection or preventing it if authentication fails. It is used for most wireless 802.11 access points and is based on the Extensible Authentication Protocol (EAP).

Through the use of the certificate embedded into the SIM, the mobile device can join a Wireless network protected via EAP-TLS, for example public hot-spots based on IEEE 802.11 a/b/g, WiMAX or future wireless protocols that use IEEE 802.1x as authentication standard.

VIII. Lawful interception

Lawful interception is required in some countries. X.509 certificates used in cryptography are an obstacle to lawful interception, therefore this factor must be taken into consideration.

Each mobile operator should generate a key-pair exclusively for lawful interception and the private key must be kept in a smartcard or in a secure device such as HSM so that wiretapping can be made possible, yet secure. The public certificate must be stored in the ESIM in a well-known location and made available to application requiring it, such as the Short Messaging Service, using common APIs.

It is a key factor that is the *home operator* owning the wiretapping certificate, not the visiting operators in which the user is roaming. This ensures privacy of communication in those countries that offers GSM or other mobile services but has a different privacy regulation. It is important to stress that the user is responsible of its content in his/her home country, while ensuring content privacy while roaming. If an international crime is committed, local government justice can ask for wiretapping following the international justice.

For the services based on SIM Tooklit, such as banking application, that are in need of the confidential information, the lawful interception should not be made available by the operator itself. Instead, the service provider should maintain logs and provide information to local government authorities upon requests.

IX. Criticism

Certification Authorities and public key infrastructure had not the expected success: it had a wide adoption on the server side for web access, but not on the client site. We are facing a low adoption for the client authentication, maybe associated to the costs involved in maintaining a PKI infrastructure and the usability issues for an end user. However smart cards and client X.509 authentication are slowly begin their adoption

where sensitive data is involved, such as government agencies and military armies.

History told us that in the past peer-to-peer cryptographic mechanism such as PGP or SSH were the preferred methodologies for security over adopting the PKI infrastructure.

On the mobile operator side, mediation through third parties comes to disasters, such as what happened to WAP and other operator-mediated initiatives.

However, I believe that the pitfalls of these systems are mainly due to costs and the "usability" factor from an user perspective. For example, WAP on one side had poor rendering of web pages, basically because there were no standards on the mobile web browsers and each one need to be an exception on the WAP gateways, on the other the costs of web surfing over mobile was too high.

From a security perspective, the user should not be aware of the underlying security technology that must be transparent to the end-user. When the user enter his/her PIN, it must unlock also the on-board certificate chains so that he will notice no difference, but he/she is having real security.

PosteMobile, the MVNO owned by Poste Italiane (italian post), filed a patent[15] that describe their methodology, based on the X.509 public key infrastructure, to secure data exchange between the SIM Toolkit Application and the webservice that interacts with the PostePay service, the pre-paid credit card service of Poste Italiane. Such a successfully example demonstrates that a public key infrastructure can be widely adopted without the actual knowledge of the underlying security technology from an user. However the system adopted by PosteMobile is limited only to the SMS messages exchange, while the ESIM proposal has the potential of addressing more security needs.

X. Acknowledges

Many thanks to Prof. Donal O'Mahony, director of CTVR at Trinity College Dublin (Ireland), that made this publication possible. Tommaso Cucinotta and Ludovic Rousseau for the precious suggestions.

References

[1] *Digital cellular telecommunications system (Phase 2+); Specification of the Subscriber Identity Module - Mobile Equipment (SIM - ME) interface (GSM 11.11)*. ETSI, 1995, vol. GSM Technical Specification.

[2] *Digital cellular telecommunications system (Phase 2+); Specification of the SIM Application Toolkit for the Subscriber Identity Module - Mobile Equipment (SIM - ME) interface (GSM 11.14)*. ETSI, 1996, vol. GSM Technical Specifications.

[3] "Iso 7816-4 smart card standard: Part 4: Interindustry commands for interchange," International Organization for Standardization (ISO), 2005.

[4] G. Paternò, "Exploring smartcards, an independent look to technologies and market," Tech. Rep., 2008. [Online]. Available: http://www.gpaterno.com/publications/2008/Exploring_Smartcards.pdf

[5] Java card technology specifications. [Online]. Available: http://java.sun.com/javacard/

[6] D. Corcoran and T. Cucinotta, *MUSCLE Cryptographic Card Edge Definition for Java Enabled Smartcards*, 2001, no. 1.2.1.

[7] T. Cucinotta, M. D. Natale, and D. Corcoran, "A protocol for programmable smart cards," in *Proceeedings of* 14^{th} *International Workshop on Database and Expert Systems Applications (DEXA 2003)*, IEEE. Prague, Czech Republic: IEEE Computer Society Press, September 2003, pp. 369–374.

[8] ——, "An open middleware for smart-cards," Computer Science Software Engineering (CSSE) Journal, 2005. [Online]. Available: http://retis.sssup.it/%7Etommaso/publications/CSSE-2005.pdf

[9] Microsoft. Windows mobile certificatestore apis description. [Online]. Available: http://msdn.microsoft.com/en-us/library/dd187703.aspx

[10] Apple. Keychain services concepts. [Online]. Available: http://developer.apple.com/DOCUMENTATION/Security/Conceptual/keychainServConcepts/02concepts/concepts.html

[11] A. Greenberg, "How to hijack 'every iphone in the world'," *Forbes*, 2009. [Online]. Available: http://www.forbes.com/2009/07/28/hackers-iphone-apple-technology-security-hackers.html

[12] M. Montoro, "Cain & abel." [Online]. Available: http://www.oxid.it/

[13] T. Kaji, K. Hoshino, T. Fujishiro, O. Takata, A. Yato, K. Takeuchi, and S. Tezuka, "Tls handshake method based on sip," *Proceedings of the International Multiconference on Computer Science and Information Technology*.

[14] M. Baugher, D. McGrew, M. Naslund, E. Carrara, and K. Norrman, "The secure real-time transport protocol (srtp)," IETF, March 2004. [Online]. Available: http://www.rfc-editor.org/rfc/rfc3711.txt

[15] F. Fucci and R. Giacchi, "System based on a sim card performing services with high security features and relative method," World Intellectual Property Organization, 2009. [Online]. Available: http://www.wipo.int/pctdb/en/fetch.jsp?LANG=ENG&DBSELECT=PCT&SERVER_TYPE=19-10&SORT=41273866-KEY&TYPE_FIELD=256&IDB=0&IDOC=1628445&C=10&ELEMENT_SET=B&RESULT=2&TOTAL=6&START=1&DISP=25&FORM=SEP-0/HITNUM,B-ENG,DP,MC,AN,PA,ABSUM-ENG&SEARCH_IA=IB2008054047&QUERY=(FP/giacchi)+