A SECURE AUTHENTICATION INFRASTRUCTURE FOR MOBILE COMMUNICATION SERVICES OVER THE INTERNET

Irénée Dupré la Tour, Gregor v. Bochmann, and Jean-Yves Chouinard
School of Information Technology and Engineering, University of Ottawa
161 Louis-Pasteur, Ottawa, Ontario, Canada K1N 6N5
{idupre,bochmann,chouinar}@site.uottawa.ca

Abstract Mobile communication on the Internet sets more security concerns than traditional mobile networks such as GSM. The network infrastructure registration process should give credentials to the user to let him or her being identified by any service provider in order to prevent fraudulent use. In addition, a user should be able to communicate with privacy and to sign a message (e.g. a payment order) so that billing is possible. Users should be able to connect from everywhere, with various types of terminals, possibly mobile. In this paper, we propose to secure an infrastructure providing telecommunication services on the Internet for a mobile user. We establish a trust relationship between any pair of the parties with a password-based user access. As for user-to-user communication, both signaling and media data can be secured. We illustrate the use of this infrastructure to provide secure IP-Telephony.

Keywords: Security, Mobility, IP-Telephony

1. INTRODUCTION

As the commercial use of the Internet becomes more common and the demand for mobility through the Internet increases, it is necessary to provide a scalable authentication infrastructure and key distribution support for multimedia communication. One of the applications requiring such authentication infrastructure is Internet Telephony. Schulzrinne explains in [1] that, while using the term of Internet Telephony, "it should be understood that the addition of other media, such as video or shared applications, does not fundamentally change the problem." Indeed, unlike the public switched telephone network (PSTN), radio or television

networks, the Internet is not an application-oriented network and the delivery of stored (streaming) music or video and telephone-style applications can share almost all of the underlying protocol infrastructure [2]. This paper proposes a scalable authentication infrastructure for Mobile Internet Telecommunication services (MobInTel). We present this infrastructure in details and illustrate its use with secure IP-telephony.

2. MOBILE COMMUNICATION INFRASTRUCTURE

In this paper, we define *User (or personal) mobility* as the ability of a user to access telecommunication services from any terminal (e.g. workstations, notebooks, Personal Digital Assistants, cellular phones) at any place in the world on the basis of a personal unique identifier, and the capability for the network to provide services in accordance with the user's service profile. *Session (or service) mobility* refers to the ability to continue a suspended session on another terminal. Users have the capability to suspend a session at one desk and pick it up elsewhere on the network. *Terminal mobility* is the ability to maintain communications while moving the terminal (e.g. cellular phone) from one sub-net to another. Terminal mobility is typically associated with wireless access.

A mobility architecture, as considered in this paper, includes all the three kinds of mobility described above. It typically involves three parties: the user (say Alice), the Home Agent (HA) and the Foreign Agent (FA). The MobInTel infrastructure [3] provides personal mobility using the home directory concept and the agent-based infrastructure. This architecture provides multimedia services with global mobility (terminal, user, and session). The Internet is divided in a large number of network domains (sub-networks). Each network includes a Service Agent that acts as HA for users registered in that domain and as a FA for other users. The Service Agent also includes a *user home directory*. This directory includes information about users registered in that domain concerning authentication, authorization, accounting, Quality of Service (QoS) preferences and location. Only the HA has access to this directory. Quality of service negotiation based on device capabilities and user QoS preferences stored in the user home directory is presented in [4]. It is an example of use of information stored in the user home directory. Information about authentication are used in the protocol we define below.

If Alice connects in her home domain, a direct trust relation can be established with the HA and the whole authentication process is

much simpler. We assume in our scenario that Alice connects to the infrastructure from a foreign domain. It is important for the FA to authenticate the user and ensure that the user is legitimate so that billing is possible. The trust relationship between Alice and the FA is based on their trust relationship with the home agent. Both Alice and the FA trust the HA during the registration process. Although an explicit authentication between the FA and the user would be better, our scheme is not unreasonable since it spares bandwidth and computational power for the mobile user terminal.

Before using the infrastructure, Alice must register in a network domain (her home domain). This means that the home agent and Alice share a security association. In our scenario, we consider the commonly found case of security association based on a shared password, due to its practicality. We assume that both the FA and the HA own a digital certificate and that Alice, at least at the beginning of the authentication process, cannot verify the validity of certificates. While we suggest the use of digital certificate-based authentication and security association establishment between FA and HA, the operations rely on mechanisms provided by other infrastructures discussed in Section 4. We take advantage of the growing public-key infrastructure (PKI) to check the validity of digital certificates. The mobile user is assumed to have limited computational power. Alice should avoid using a public key algorithm as much as possible since most public-key algorithms tend to be computationally intensive. The home agent and the foreign agent are assumed to have enough computational power to perform public key encryption and certificate signing. Alice may communicate with the foreign agent via a wireless link. Such links are particularly vulnerable to passive eavesdropping, active replay attacks, and other active attacks. The security requirements include protection against fraud, efficiency (in term of computational complexity and required bandwidth), distributed management and confidentiality of user identity.

At the end of the authentication process, Alice shares two new security associations: one with the FA and one with the HA. She also has a digital certificate so that everyone who can check the validity of such certificate can authenticate her. Once authenticated, Alice can use services provided by the FA. Furthermore, she can make any purchase on the Internet using her certificate to sign a payment order. This would require to use a computationally intensive algorithm but only on a small amount of data. Say Alice wants to call Bob. She contacts the IP telephony server in Bob's home domain to transmit the call request. She contacts the bandwidth broker in her current domain (FA's domain) to reserve resources for the phone call. The bandwidth broker can bill Al-

ice using her signed payment information. Bob and other new parties can authenticate Alice using her digital certificate. Our authentication framework is not linked to a particular local service provider and thus it could be used to support any service provider.

3. EXISTING INFRASTRUCTURES AND PROTOCOLS

Most existing mobile systems, such as the Global System for Mobile Communications (GSM), do not transmit all communications on the Internet, and thus lead to different security requirements. GSM provides terminal mobility only and it is based on a fixed signaling network that is assumed to be secure. In such homogeneous mobile user environments, no operations between the foreign domain and the home domain are needed, or these operations are static (e.g. roaming agreements). However, the Internet is formed by a set of heterogeneous networks, administrated locally. No trust relationship exists between a home domain and a foreign domain before they authenticate each other. As a result, the approach taken by GSM cannot be simply transposed to the Internet environment. Moreover, it is not scalable to consider defining security associations between pairs of foreign and home agents. A centralized key distribution center (KDC) is used in Kerberos [5] to assist authentication and key management. In the Internet, it is very common that a long distance exists between KDC and the foreign/home domain, and thus long delays are introduced in communication with the KDC. A reasonable authentication and key distribution scheme should be managed on a distributed, rather than centralized basis, since the application environment is entirely distributed. These observations strongly suggest that we take the public key approach for designing an authentication and key distribution scheme.

Telephony on the Internet means that both signaling and communication data are transmitted through the Internet. General public Internet Telephony products are currently not secured [6]. Some telephony software introduced various kinds of security features but no architecture takes into account both QoS and security requirements. In the latest version of Microsoft Netmeeting^{TM}, only user authentication and data encryption (excluding audio and video) are provided. PGPfone^{TM} makes use of a biometric signature scheme based on voice to authenticate users but this scheme is not completely reliable and is not convenient for the user. However, in the latter case data encryption can be provided. There are two main telephony-signaling protocols on the Internet: one defined by ITU (International Telecommunication Union) within H.323 [7] and

SIP (Session Initiation Protocol) [8] defined by the IETF (Internet Engineering Task Force). H.323 is a set of recommendations, which defines how voice, data, and video traffic will be transported over IP-based local area networks and the Internet. SIP is an application-layer control protocol for creating and terminating sessions such as Internet telephone calls. SIP in itself supports user mobility by redirecting requests to the user's current location. Users can register their current location. SIP supports user location, device capabilities, user availability, call setup and call handling.

4. SECURED AUTHENTICATION PROTOCOL

The proposed protocol uses a broadcast message and a 2-way authentication process. The broadcast message informs the user about the FA location. This kind of message is necessary for any mobile user connecting to a foreign domain. Fig. 1 illustrates the message exchange sequence for user authentication. Messages $\overline{M_4}$ and $\overline{M_5}$ are sent in case of a negative authentication answer (otherwise messages M_4 and M_5 are sent).

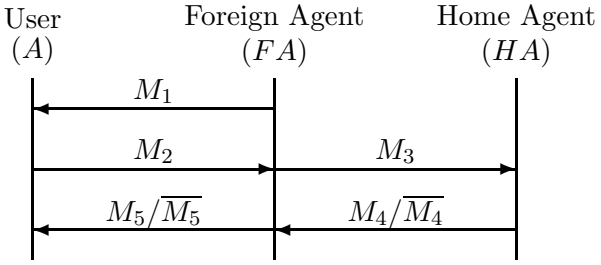

Figure 1 Authentication protocol sequence diagram.

Broadcast information: $M_1 = KU_{FA}$, "FA location"

A broadcast message (through a dedicated advertisement agent or an existing agent such as DHCP or Mobile IP) in the local domain informs Alice about the location and the public key KU_{FA} of the MobInTel agent.

Authentication request: $M_2 = KU_{FA}(Ks_1), Ks_1(ID, KU_A, N_1, DP, HV)$

Alice picks a random session key Ks_1 that will be used only for message to avoid encrypting the whole message with FA's public key. The message includes Alice's identity (e.g. *alice@domain.net*) and home domain address, so that FA knows in which domain to forward the authen-

tication request. For the digital certificate request, Alice generates a pair of public and private keys KU_A and KR_{Alice} on her terminal. KR_{Alice} is stored on the terminal in a secure way and is never sent on the network. She sends KU_A with the authentication request so that HA can bind Alice's name and KU_A. In other words, Alice sends a Certificate Signing Request (CSR) to HA that produces a digital certificate. HA acts as a Certificate Authority (CA) and manages the process of issuing, renewing, and revoking certificates. HA may be just one branch of the certification tree so that HA's authority can be signed by a higher level authority. A nonce N_1 is sent for key management, to guarantee the integrity of previous parameters and to avoid certain types of attacks. The device profile DP is sent to let HA know which type of media this device supports. The hash-value $HV = H(ID, N_1, DN, DP, KU_A, pwd)$, where $H()$ is a one-way function, contains information that allows HA to authenticate Alice, for instance, it may be obtained by hashing Alice's ID and her password. DN (domain name), DP and KU_A are included in the input of the hash function to guarantee their integrity.

Authentication request forward: $M_3 = SecCx(ID, KU_A, N_1, DP, HV$

Before forwarding the request, FA keeps track of certain parameters: Alice's ID, Ks_1 (current session key with Alice), N_1 that will be used to create the new session key with Alice, and DP that gives FA information about the type of media Alice can receive on the device she is connecting from. This message is sent to Alice's HA. The FA must be able to retrieve HA's location knowing only the name of the domain. This could be done with a DNS lookup in HA's domain. Finally, when FA knows HA's location, it sends the message over a secure connection.

Authentication reply (ACK/NACK): $M_4/\overline{M_4}$

$M_4 = SecCx(ID, ACK, HV, Ks_2, N_2, N_3, CERT_A)$

$\overline{M_4} = SecCx(ID, NACK, HV, HV2)$

Using the current secure connection ($SecCx$) established with FA, HA sends back the answer including Alice's ID, the answer of the authentication process (ACK), the hash-value sent by Alice that uniquely identifies the request, Ks_2 and the session-key that will be used between Alice and FA. The nonces N_2 and N_3 will be forwarded to Alice and are used to calculate Ks_2 and Ks_3 knowing N_1 and Alice's password (pwd). The FA receives Ks_2 in clear over the secure connection with the HA. In case Alice is not authenticated, that is $HV_{HA} \neq HV$ (HV_{HA} being the hash-value calculated by HA), the authentication reply message includes a $NACK$ (negative acknowledgement with possibly a reason e.g. "revoked user"), previous values to identify the request (ID and HV) and an additional value $HV2$. $HV2 = H(pwd, N_2, HV)$ is the digest of HV, nonce N_2 and Alice's password. $HV2$ will be sent to Alice as a proof

that $NACK$ is the answer from HA and that FA has communicated with FA to get the answer. The nonce N_2 prevents a cryptanalyst to perform a chosen plaintext attack on the password given the pair $(HV, HV2)$.
Authentication reply forward (ACK/NACK): $M_5/\overline{M_5}$
$\overline{M_5} = Ks_2(ID, ACK, HV, N_3, CERT_A), N_2$
$\overline{M_5} = Ks_1(ID, NACK, HV, HV2)$

Alice computes $Ks_2 = H(N_1, N_2, pwd)$ and then tries to decipher $Ks_2(ID, ACK, HV, N_3, CERT_A)$. If she succeeds, Alice knows that FA received the key from HA. That means FA communicated with HA and was authenticated as a valid agent (HA checked FA's certificate). Finally, she computes $Ks_3 = H(N_1, N_3, pwd)$, the session key to be used between Alice and HA during the session. If not acknowledged, FA answers to Alice using Ks_1. This message indicates the authentication failure and the authentication identifier (ID and HV). $HV2$ is sent to Alice as a proof that $NACK$ is the answer from HA and that FA has communicated with FA to get the answer. Indeed only HA could have generated $HV2$.

The cryptographic hash function used in the authentication protocol may be keyed SHA-1 (or possibly keyed MD5) with a key size of 160 bits (respectively 128 bits). The password is used as a key that can be filled to reach the required size using the same pad defined for the considered hash algorithm. Private-key algorithms should be chosen such that the length of the key can be adapted to the computational power of the user terminal. AES (Advanced Encryption Standard) and Blowfish are such algorithms. Elliptic Curve Cryptography (ECC) should preferably be used for public-key encryption rather than RSA to make use of its shorter key length at equal security level. Secure connections could be set up in several ways since both FA and HA own a digital certificate. TLS (Transport Layer Security), IPsec (IP security), IKE (Internet Key Exchange) or any secure link establishment protocol could be used between the two agents. Messages could be formatted in XML to combine simplicity and compatibility with other protocols and standards.

5. USING THE AUTHENTICATION INFRASTRUCTURE FOR IP TELEPHONY

Alice does not trust the HA to establish secure communication with parties other than FA. Thus she can establish a session key with Bob without HA knowing it. Alice should also be able to phone Bob with privacy and anonymity. To provide the latter, signaling messages should be encrypted all along the way.

Let us see the scenario of Alice calling Bob using SIP. A successful SIP invitation consists of an $INVITE$ request (call request) followed by a

response from the callee and an ACK (acknowledgment) from the caller. Alice can either send the invitation request to a local SIP proxy, or send it directly to the callee. She can find a SIP server by querying a DNS. The SIP server of the callee (Bob) can act as a proxy server (forward call $INVITE$ to Bob) or as a SIP redirect server (that sends back the user location to the caller). We suppose below that it is a proxy server and it forwards the call $INVITE$ to Bob's location. The response message takes the reverse path to reach Alice. Alice then sends an ACK message using the same path.

A SIP message can be divided into two parts. The first part contains a start-line and some fields of the header that have to remain in clear (including the identity of the caller and the callee) for various reasons. The second part contains other fields of the header that can be encrypted and the body. SIP defines some built-in security features. SIP message authentication can be provided by strong signature. Built-in SIP encryption schemes provides encryption (using PGP or another scheme) of the second part of the SIP message only. If the requirements of the architecture include privacy of caller and callee identities, a lower layer security protocol must be used to encapsulate SIP (e.g., IPsec, TLS or another protocol). QoS capabilities exchange can be done through fields carried in the SIP body. Once the connection parameters are known, resource reservation (using the resource reservation setup protocol) can be done using local bandwidth brokers. Then media streams can be sent using the Real-Time Protocol (RTP).

Fig. 2 shows the message exchange sequence for Alice calling Bob with SIP. We assume that both Alice and Bob have registered in their respective foreign domain SIP servers. Registration in a foreign domain requires two register messages. That can be done securely assuming each SIP server has a certificate. Before sending an $INVITE$ message to Bob, Alice must ask Bob's home agent for Bob's certificate. This is necessary to provide end-to-end encryption of sensitive data. All SIP messages can be totally encrypted using TLS or IPsec. Since each party has a certificate, these secure link establishments are possible. For the $INVITE$ message, M_1, the second part of the SIP message is encrypted with Bob's public-key and signed by Alice with her private-key. This message also includes Alice's certificate obtained during the authentication phase. The second SIP message (M_2) is the message forwarded by the SIP server in Bob's home domain to the one in Bob's foreign domain. The same process is done by the SIP server in Bob's foreign domain (M_3). Then Bob verifies the signature with Alice's certificate and decrypts the second part of the SIP message with his private key. For the reply, Bob encrypts the same fields with Alice's public-key and

signs these two parts with his private-key M_4. This message is then transmitted to Alice using the reverse path (M_5 and M_6). Encryption and signature on the second part of the header and the body is end-to-end. The ACK messages, M_7, M_8, and M_9 are encrypted and signed the same way as the first three messages (without sending Alice's certificate). They contain the ACK response information related to the call.

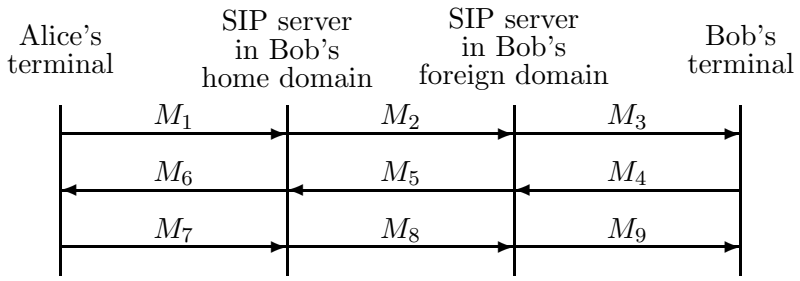

Figure 2 Phone call use case: message exchanges.

6. SECURITY ANALYSIS

Some users may require anonymity or location privacy. In order to acquire location privacy, the user name is encrypted with the public key of the foreign agent. This does provide user location privacy. A unique alias to replace the real user identity could be used, but once the static mapping between real identity and alias is disclosed, the user location will be exposed as well. Billing is the keystone of commercial use. When the foreign agent (or a local service provider) sends billing information to the home agent, it may use a similar scheme as the one of Secure Electronic Transaction (SET). The user's dual signature related to the purchase could be sent to both FA and HA (acting as a payment gateway) so that FA doesn't know about the payment information and HA doesn't know about the service Alice is asking to pay for. This way, purchase privacy is guaranteed. It should also be noted that a good user password choice is essential. The home agent should prevent the user from keeping a weak password that can be guessed or found easily. These passwords should be identified before they are broken by constantly running password cracking programs. In our scheme, even the hash-value (HV) is encrypted. This technique increases the computational overhead of cracking passwords as advocated in [9].

Let us study different kinds of attacks against the protocol proposal and how they are addressed. First, consider *spoofing attacks*: a malicious user, say user Z, may try to usurp Alice's identity. Authentication information is included in the value of HV sent by Alice to the foreign agent in M_2. Since Z does not know Alice's password, the HA while calculating HV_{HA} will find a different value and will not authenticate Z as Alice. In M_4, HA sends the authentication result to FA so that FA knows that Alice (actually Z) is not authenticated. Spoofing of servers (FA, HA, SIP servers) is denied by the systematic use of digital certificates. In the same way, Alice and Bob authenticates Alice with their certificates that provide end-to-end encryption on sensitive parts of the message. *Replay attacks* are impossible owing to the nonces. If an attacker tries to replay M_2, this will be detected by HA that keeps all successful login nonces for a given time (e.g. a few days). Since the nonce N_1 includes the date, this prevents any replays. Another way to do it would be to ask Alice to send a confirmation message to HA as a seventh message saying that she has decrypted message M_5. SIP messages cannot be replayed if secure connections between Alice, Bob and the different servers include replay attack prevention such as in TLS. *Denial of service attacks* (a.k.a. "DoS") are possible since each authentication request consumes both bandwidth and processing time for FA and HA. This is a general issue for any service on the Internet. This can be avoided by using adaptive firewalls or intrusion/attack detector systems [10]. DoS attacks are made easier since each INVITE message requires some computation. This is an inevitable trade-off between efficiency and security. In [8, §13.4], the authors underline that unauthenticated reply messages should be ignored since they could be sent by a rogue proxy if link-by-link encryption and authentication is not systematically chosen.

7. CONCLUSION

In this paper, we have proposed a secured authentication infrastructure for mobile communication over the Internet. The authentication is based on a secret password, which is also known by the user's home agent. The characteristics of Internet communication are taken into account. In general, the mobile user first talks to a foreign agent, which in turn communicates with the home agent. The essentials of the protocol are summarized as follows. The foreign and home agents authenticate each other with certificates. The user and the foreign agent authenticate each other through the home agent that is trusted by both. At the end of the authentication process, the user gets a terminal-specific certificate that allows him to sign and thus to authenticate a key exchange or

communication request with another user. The home agent proposes the session keys between the user and the agents. These keys can be used in subsequent communication between the user and the other agents. Sensitive information such as session keys and authentication information are always encrypted during the exchanges. We showed that this mechanism works for IP-telephony using SIP and, therefore, this scheme can be used to provide any mobile application over the Internet for various service providers and especially multimedia communications.

REFERENCES

[1] H. Schulzrinne and J. Rosenberg, *Internet Telephony: architecture and protocols – an IETF perspective*, Computer Networks, vol. 31, No. 3. February 11, 1999.

[2] C. A. Polyzois, K. H. Purdy, P. Yang, D. Shrader, H. Sinnreich, F. Ménard, and H. Schulzrinne, *From POTS to PANS – A Commentary on the Evolution to Internet Telephony*, IEEE Network, vol. 13, no. 3, pp. 58–64, May/June 1999.

[3] X. He, K. El-Khatib, and G. v. Bochmann, *A communication services infrastructure including home directory agents*, Technical report, University of Ottawa, Canada, May 2000.

[4] X. He, K. El-Khatib, and G. v. Bochmann, *Quality of service negotiation based on device capabilities and user preferences*, Technical report, University of Ottawa, Canada, May 2000.

[5] J. Kohl and C. Neuman, *The Kerberos Network Authentication Service (V5)*, RFC 1510, IETF, September 1993.

[6] C. Rensing, U. Roedig, R. Ackermann, and R. Steinmetz, *A Survey of Requirements and Standardization Efforts for IP-Telephony-Security*, Darmstadt University of Technology, Germany, Proceedings of the Workshop "Sicherheit in Mediendaten", September 2000.

[7] ITU-T Recommendation H.323 V.3, *Packet-Based Multimedia Communication Systems*, September 1999.

[8] M. Handley, H. Schulzrinne, E. Schooler, and J. Rosenberg, *SIP: Session Initiation Protocol*, RFC 2543, March 1999.

[9] R. Dhamija and A. Perrig, *Déjà Vu: A User Study using Images for Authentication*, 9th Usenix Security Symposium, August 2000.

[10] W. Stallings, *Cryptography and Network Security: Principles and Practice*, 2nd ed., Prentice-Hall, 1999.