Authentication Protocols for Personal Communication Systems*

Hung-Yu Lin and Lein Harn
Computer Science Telecommunications Program
University of Missouri-Kansas City
4747 Troost
Kansas City, MO 64110

ABSTRACT

Masquerading and eavesdropping are major threats to the security of wireless communications. To provide proper protection for the communication of the wireless link, contents of the communication should be enciphered and mutual authentication should be conducted between the subscriber and the serving network. Several protocols have been proposed by standards bodies and independent researchers in recent years to counteract these threats. However, the strength of these protocols is usually weakened in the roaming environment where the security breach of a visited network could lead to persistent damages to subscribers who visit. The subscriber's identity is not well protected in most protocols, and appropriate mechanisms solving disputes on roaming bills are not supported either. To solve these problems, new authentication protocols are proposed in this paper with new security features that have not been fully explored before.

I. INTRODUCTION

In personal communication systems, open access to the radio exposes the context of communication over the wireless link between a mobile unit and the wired network. Such openness also gives an intruder the opportunity to masquerade as a legitimate subscriber to make free calls. To provide proper protection on this wireless link, security features, such as confidentiality and fraud control, need to be provided. In principle, these features can be achieved through authentication protocols that verify the identities of entities on both ends of the wireless link and establish a secret session key between them for the following secret communication. Although, protocols on wired networks with similar features have been available, it woul not be appropriate to apply them directly in the PCSs environment because of some specific requirements in the wireless environment. For example, considerations on hardware complexity, battery power, and validation delay have forbidden a mobile unit from performing computations that require expensive hardware or are time-consuming (i.e., high power consumption). Recently, several authentication protocols for PCSs have been proposed by standard bodies [6, 9, 10, 11] and independent researchers [1, 2, 5, 12, 13]. With different considerations in mind and techniques used, each one has its own pros and cons on different applications [3, 5, 7, 13, 17, 18].

The service of personal communication systems is provided by multiple regional networks, each operated under a different administration. One subscriber could roams among several networks. For most systems, the subscriber and his home network share an authentication key with which they can prove themselves to each other during the authentication process. In the roaming situation, instead of the authentication key itself, some security parameters derived from the authentication key are sent from the roamer's home network to the visited network so it can perform the authentication process. To minimize the delay caused by the interactions with the home network during the authentication process, either several sets of security parameters are generated and transferred in batch to the visited network in advance of the authentication, or the same values of the security parameters are repeatedly used in several instances of authentication to cut some traffic. Without transferring the ultimate secret (the authentication key) to the visited network, this approach reduces the risk of exposing the authentication key, which causes serious damage to the service. In case of a security breach, the security of the service is expected to recover after the compromised security parameters expire. The security can also be recovered by discarding these security parameters if such compromise is detected. Nevertheless, these security parameters impose extra security burden on the visited network for its storage and management. As there are many networks in the PCSs, each operated under a different administration with a different level of protection, some networks are more vulnerable than others to attacks from intruders or insiders. Once these security parameters are compromised, either by an intruder or an insider, fraud will happen in the designated period. Sometimes, the damage caused is far more serious and persistent than what is first thought. For example, if the (Kc, R, S) tuple in GSM [11], KS in DECT [10], or the SSD in USDC [9] are known by the attacker, he can use it to impersonate a legal network (base station) to establish a connection with the corresponding subscriber, and hence can draw some private information of the conversation, e.g., the identities of parties involved in the conversation. The attacker can also then pretend to be the other party of the call to gather further information until such impersonation is detected. This attack can be repeatedly launched in GSM because the mobile unit is not designed to detect used security parameters. In USDC, though, a new SSD

* To appear on SIGCOMM'95

can be re-established through SSD Update Protocol. Unfortunately, it can only be invoked by the serving network. If an intruder tries to masquerade as a legal network, he won't invoke SSD Update Protocol at all. In DECT, an independent protocol is provided for the subscriber to verify the serving network. Invoking this protocol is optional. If a subscriber suspects that the serving network is being impersonated, he can invoke this protocol at the cost of either extra delay or bandwidth. Even with public-key approach, some subscriber-specific sensitive information could also be found in a visited network, e.g., the common key η of RCE and MU in MSR+DH [2]. With knowledge of η, the attacker can always impersonate the legal subscriber within this specific network.

Under some situations, an old session key can also be derived without having to break a mobile unit or a service provider. With this compromised session key together with other recorded information, an attacker can make fraudulent calls or masquerade as a legitimate serving network to establish a false connection with the subscriber.

In this paper, new authentication protocols are developed to provide better security for the personal communication systems. These protocols will solve problems resulting from compromised session keys or security breaches on weak visited networks. New security features, which include subscriber-ID confidentiality and the mechanism to solve disputes over roaming bill, are also provided. One distinct feature of these protocols is the use of conventional secret-key techniques in combination with modern public-key techniques, and the trade-off made between these two techniques. For the rest of the paper, we will first outline the desired security features and implementation requirements of the authentication protocols for PCSs in Section 2. Then, new protocols are proposed in Section III along with the analysis in Section IV. A short conclusion is given in Section V.

II. DESIRED SECURITY FEATURES AND IMPLEMENTATION REQUIREMENTS

A. Security

1. Session key establishment: Radio signals transmitted over the air in current cordless and cellular systems can be intercepted easily by commercially available scanners. In the advanced digital systems, this problem still exists. To protect the contents of communication from eavesdropping, messages must be transmitted in ciphertext. Therefore, during the authentication process, a common secret must be agreed upon by the subscriber and the serving network. This session key can be repeatedly used in some situations as found in [13] and optionally in DECT [10]. But due to security concerns, most protocols require a new key for each session.

2. Caller ID confidentiality: In the traditional telephone system, a subscriber is connected to the local office through a fixed line. This line automatically identifies the subscriber (the phone number). However, in this wireless environment, without such physical association, a subscriber has to somehow provide his identity to the serving network for necessary verification. As a subscriber's identity, i.e., his current location, may be of special value to some persons [16], the actual identity of the subscriber should not be exposed to the outsider. Unfortunately, such confidentiality is not rigorously supported in current standards. In fact, a roamer's identity can even be hidden from visited networks.

3. Mutual authentication: In earlier cellular systems, a call request made by a roamer is granted even while authentication is still underway. By the time the result comes out, several fraudulent calls may have already been completed. Such delay is because of the lack of proper intercarrier communication and has caused billions of dollars in losses to the carriers [8]. Though with the establishment of intercarrier agreement, the validation process can be completed before the first call is granted, i.e., the so called first-call shut down, the modification of serial numbers and the eavesdropping on radio signals still leave the door open for wicked attackers to commit fraud. With the emergence of new digital systems, modern cryptographic techniques can now be used to eliminate such fraud caused by masquerading. A similar problem is the impersonation of a serving network by the intruder, which causes problems as mentioned in Section I. Therefore, it is important for a subscriber and the serving network to mutually authenticate each other in the authentication process.

4. Non-repudiation of service: For the service provider, it is desirable that a subscriber cannot deny the bill incurred from services he requested. Similarly, the subscriber should not be wrongly charged due to any billing error or security breach on the serving network. Theoretically, both goals can be achieved through the use of digital signatures [14]. But it is never recommended in any standard for the large amount of computations involved. In this paper, we will try to provide a limited version of the non-repudiation service between roamers and the visited network.

B. Implementation requirements

An important consideration to provide proper authentication over this wireless link is the computational overhead on the mobile units. Because of the considerations about hardware complexity, battery power, and computation delay, some mobile units, e.g., pocket cellular telephones, cannot perform complicated operations that require expensive hardware or are time-consuming. Such limitations almost exclude the use of, generally, time consuming public-key cryptographic techniques that can provide the desired non-repudiation service. Probably, this is why strong subscriber-ID confidentiality is not supported in current standards. Cellular Digital Packet Data [6], MSR+DH [2], and the one proposed by Aziz and Diffie [1] are examples that use public-key techniques. These protocols require many more computations in comparison with those that use one-key cryptographic techniques. Another concern with public-key approaches is the revoking of certificates, which require complicated mechanisms to handle. Conventional one-key cryptographic algorithms provide fast operations on enciphering and deciphering, and are therefore used for secret communication if one common secret key is agreed upon among the communicating parties. Even in the authentication process that establishes a common session key for this wireless link, they are still preferred in current standards.

Another consideration is the validation delay. During the validation process, in addition to the computation performed by the participating parties, messages are exchanged between a subscriber and the serving network or between a visited network and the subscriber's home network. These interactions, especially those on the radio, cause validation delay. Reducing such interactions is an important issue in designing authentication protocols.

III. PROTOCOLS

A. Motivation

Though a subscriber does not need to share a secret authentication key with the serving network as demonstrated in protocols [1, 2, 6] using public-key techniques, these protocols inevitably involve a great deal of computation and the key certification/revocation problem. This is why protocols proposed here choose to have a secret authentication key shared between a subscriber and the serving network. However, to reduce a roamer's trust on the visited network's capability of protecting sensitive data (which is specific to the roamer) so that enhanced security services can be provided, public-key techniques are also incorporated in the proposed protocols. But note that it is the network operator only, not the subscribers, who needs to choose a private key and then makes the corresponding public key known to its subscribers according to the chosen public-key cryptosystem. This approach avoids the problems of key certification/revocation. Complications of implementation and computations incurred can also be absorbed by the network, which has sufficient computing resources, instead of the subscriber, who may only be equipped with a pocket-sized device.

B. Terms and notations

The service of personal communication systems is provided by multiple regional networks, each operated under a different administration. Every network chooses its own public-key cryptosystem. When a user wants to subscribe to the service, he chooses one network as his home network (HN) and becomes home subscriber of this network. Upon subscription, the subscriber gets a secret authentication key k shared with his home network and the home network's public key e_{HN}. When a subscriber roams into an area operated by another network, he has to register at this visited network (VN) and becomes a visiting subscriber of that network. A serving network, which may be a visited network or the subscriber's home network, is the one which is currently providing the service to the subscriber who may be a caller or a callee. The wireless link to be protected can be the one between the caller and the network or the one between the network and the callee. Although protocols for the latter are seldom discussed, they may not be systematically the same as the protocols for the former one. In each of the following protocols, the subscriber and the serving network will verify the identity of each other, and a new secret key will be established between them if both identities are correct. In this paper, a subscriber and the mobile unit are regarded as an intact part. Authentication between a user and the mobile unit is not covered. Following is a list of notations used in these protocols.

- [...]: a sequence of values to be encrypted
- $f_1(e; [...])$, $f_2(e; [...])$: public-key encryption functions with e as the public key.
- $g_1(k; [...])$, $g_2(k; [...])$: Secret-key encryption/decryption algorithms with k as the secret key.
- $h_1(a, b)$, $h_2(a, b)$: one-way hash functions with two parameters, or one-key encryption/decryption algorithms with a as the key and b as the message to be encrypted.
- $h_3(r)$: one-way hash function with one parameter.
- the subscriber's authentication key, which is only known by the subscriber and his home network.
- j: local counters kept in the mobile unit and the serving network, respectively.
- *UID*: subscriber's identity.
- *NID*: network identity.
- *TID, TID'*: the subscriber's temporary identity given by the serving network. It hides the subscriber's real identity from outsiders.
- k_i: session key established among the subscriber, his HN, and the VN in round I.

Note that g_1 and g_2 need not be different. Different notations are used to differentiate their roles in the protocols. For the same reason, f_1 and f_2 can also be the same. The presentation of the following protocols assumes that the subscriber is in a visited network. f_1 and g_1 are chosen by the home network, while f_2 and g_2 are universally agreed among all networks. If the subscriber happens to be in his home network, the role of VN is indeed taken by the HN, and the interaction between a VN and the HN does not exist. But such a difference is transparent to the subscriber.

C. Mobile station registrations

The following protocol is invoked when a subscriber roams into a new service area and asks for the registration.

Step 0. The base station (of the serving network) generates and broadcasts a new random number a for the next incoming call.

Step 1. When a subscriber roams into a new service area, he chooses a random number b and computes $x = g_1(k; [a])$, and $y = f_1(e_{\text{HN}}; [UID, x, b])$. He then sends y, a , and the corresponding HN's identity, *NID*, to the VN for the registration.

Step 2. The VN passes the request to HN. To eliminate replayed messages, the VN can reject the request that does not contain the fresh a.

Step 3. Upon receiving the request, the HN decrypts y to get the subscriber's ID, then fetches the subscriber's secret key k, and decrypts x to see if a is present. If so, the HN recognizes the caller as a legitimate subscriber; otherwise, the call is rejected. The HN now computes $r_0=h_1(x, b)$, $k_0 = h_2(k, r_0)$, and the common session key $k_s = h_2(k_{s-1}, r_s)$, $s=1, .., m$, where $r_t = h_1(k, k_{t-1})$, $t=1, .., m$, to be used by this registered subscriber for the following calls. At last, the HN sends k_0, b, and $c_t = h_3(r_t)$, $t = 1, .., m$, to the VN.

Step 4. VN assigns a temporary identity, *TID*, to this subscriber, sets its local variable *j* to *1*, and sends $z = g_2(k_0; [TID, b, e_{VN}])$ back to the mobile unit. This *TID* will be used later in the protocol for mobile terminations.

The subscriber, knowing *a*, *b*, and *k*, computes k_0 and decrypts *z* to see if *b* is present. If so, he believes that he has successfully registered in the legitimate VN. He now sets his local variable *i* to *1*. From now on, the subscriber and the VN will use the information gathered in this phase to mutually authenticate each other for each call of mobile station originations and mobile station terminations without interacting with the HN. In fact, if a call is to be established right after the registration, k_0 can be immediately used as the session key.

As in GSM, DECT, and USDC, the protection on the communication between VN and HN is left unspecified. The reason is obvious - the protection on the communication between VN and HN should be handled by existing inter-domain protocols. It would be inefficient if protection mechanisms between VN and HN must be invoked by each individual request from roamers. Given the protection is embedded in the end-to-end protocol between a roamer and his HN (i.e., the VN cannot read the context between the roamer and his HN), it could be redundant because the authentication process still needs to be carried out between this VN and the roamer's HN and the required protection can be easily achieved through this authentication process.

D. Mobile station originations

Within the designated period after the successful registration, each call request i made by the subscriber should go through the following steps:

Step 1. The subscriber computes $r_i = h_1(k, k_{i-1})$ and $\alpha = f_2(e_{VN}; [UID, r_i])$, and sends α to the visited network.

Step 2. The visited network decrypts α to get *UID* and r_i. If *UID* is a registered subscriber and $h_3(r_i)$ equals to the check value, c_j, it accepts this as a valid call request; otherwise, it rejects the request. For the valid request, it sets session key $k_c = k_j = h_2(k_{j-1}, r_i)$, and sends back $\beta = g_2(k_j; [r_i])$ to the subscriber. Lastly, it increases the value of *j* by *1*.

After receiving β, the subscriber decrypts it to see if r_i is present. If so, he believes he has established the common secret k_i with the legitimate network and sets $k_c = k_i$ as the session key. Finally, he increases the value of *i* by *1*.

E. Mobile station terminations

Step 1. The visited network broadcasts *TID*.

Step 2. The subscriber *TID* uses the previous session key, k_{i-1}, to compute $r_i = h_1(k, k_{i-1})$ and $\alpha = f_2(e_{VN}; [UID, r_i])$ and sends α to the visited network.

Step 3. The visited network decrypts α to get *UID* and r_i. If *UID* is a registered subscriber and $h_3(r_i)$ equals c_j, it accepts this as a valid response to receive the call; otherwise, it stops the establishment of the connection. For the valid response, a new temporary identity *TID'* is assigned to the subscriber becaused the original TID is exposed. Then it sets session key $k_c=k_j = h_2(k_{j-1}, r_i)$ and sends back $\beta = g_2(k_j; [TID', r_i])$ to the subscriber. Lastly, it increases the value of *j* by *1*.

After receiving β, the subscriber decrypts it to see if r_i is present. If so, he has established the session key $k_c=k_i$ with the legitimate network. The subscriber then stores the *TID'* as the new temporary identity and increase the value of *i* by *1*.

IV. ANALYSIS

A. Security

In the protocols proposed here, a subscriber's secret key is not transferred to the visited network. Upon registration, all security parameters, which include the challenge, response, and the session keys, required for the future authentication of the subscriber and the serving network are mutually decided by *a*, *b*, and the subscriber's authentication key, *k*. With possession of *k*, the subscriber can prove himself by presenting r_i's to the serving network. Knowing the checking values c_i's, the serving network can verify the legitimacy of the subscriber.

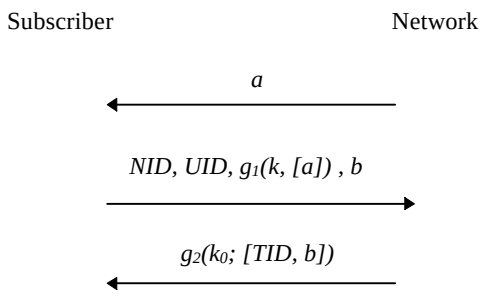
While roaming, a visited network cannot know the next session key until a visiting subscriber makes a request by presenting the correct r_i. Possible exposure of the roamer's security parameters stored in the visited network as found in GSM, DECT, and USDC is thus eliminated. Even when a (used) session key is compromised, it does not lead to the masquerading of a serving network or the masquerading of a subscriber because both entities are mutually verified for each service provided. Neither does it lead to the compromise of the next session key because r_i is not yet available, or r_i is just transmitted in ciphertext under the visited network's public key. Through such arrangements, we reduce the roamer's trust on a visited network's capability of protecting roamer-related sensitive data .

Another good feature that comes with such arrangement is the resolution of disputes on bill caused by roaming. Whenever services are provided to a roamer from foreign domains, the corresponding r_i's presented by the roamer are stored in the visited network. The visited network has no ability to compute r_i's by itself. So, under the assumption that network operators do not conspire, the visited network cannot charge a subscriber for services he does not request because the visited network will not have the correct r_i's. For the same reason, if a visiting subscriber has indeed been provided with the services, he cannot repudiate them later because, except for himself and his home network, no one else could have presented these r_i's. Note that such dispute resolution applies to all services requested, except the last one, after the registration. For the last service, the visiting subscriber can always claim that the communication is disrupted after he submits r_i, though practically, such disruption cannot occur often.

With the incorporation of public-key algorithms f_1 and f_2, subscriber-ID confidentiality is provided in these protocols. Different from previous works [1, 2] in which roamers directly verify certificates presented by visited networks, the e_{VN} accompanied by the correct *b* is deemed valid because only the legitimate network can obtain *b* from the roamer's home

network. So when a network is broken, only other networks, instead of all their subscribers, need to be notified of the voided e_{VN} (i.e., the compromised network). Roamers are free of the problem of certificate revocation.

As demonstrated in these protocols, the mutual authentication between a subscriber and the network is based on the possession of the secret authentication key. One may argue that the incorporation of public-key algorithms, which in some sense seem to be redundant, could mask the errors of the protocols. Here we can show that this worry is unnecessary. Suppose f_1 does not function at all. The protocol of mobile station registrations can be rewritten as follows:



.

Although the feature of subscriber ID confidentiality is lost, the mutual authentication is still effectively performed and the security (achieving a secret session key known only to the legitimate subscriber and the network) of the resulting protocol can be formally proved using BAN logic [4]. For the situations of mobile station registrations and mobile station terminations, the protocols are secure unless the attacker can compromise a session key and at the same time break the public-key algorithm f_2.

So far, the subscriber-ID confidentiality denotes that the subscriber's identity (i.e., his current location) is not exposed to eavesdroppers. With slight modifications to the above protocols, a roamer's (either caller's or callee's) identity can be kept from visited networks if the *TID* is given by the roamer's home network.

B. Performance

In the protocol for mobile station registrations, it takes only one round of message exchange between the subscriber and the visited network, and one round of message exchange between the visited network and the corresponding home network. For both mobile originations and mobile terminations protocols, . there is no need for the VN to contact with the roamer's HN because VN can verify the identity of the roamer and set up the session key from information given by the roamer's HN when the roamer registered. In the protocol for mobile station originations, only one round of message exchange is required between the subscriber and the serving network. While in the protocol for mobile station terminations, one extra message is required, which is inevitable, from the network to the subscriber for the notification of an incoming call. In all three protocols, the most significant computation required on the mobile unit is the operation of encryptions with public key e_{HN} and e_{VN}. Now if we choose both f_1 and f_2 to be the low exponent RSA algorithm, for example, $e_{HN}=e_{VN}=3$, then the encryption takes only two modular multiplications. For the normal *512*-bit RSA encryption/decryption operation, it takes *768* modular multiplications on average. That is, time required for the encryption of the low exponent RSA algorithm is only *1/384* of that required for normal RSA operations. (With the fastest RSA chip [15], a normal encryption/decryption operation takes 10^{-3} second.) Altogether, the minimization of interactions and simplification of computation on the mobile unit speed up the verification and therefore reduce the total delay on the authentication process.

V. CONCLUSION

Security and implementation requirements for personal communication systems have been discussed. To provide better protection, new protocols with more security features, which reduce the roamer's trust on a visited network's capability of protecting roamer-related sensitive data without involving complicated computations, were proposed and then analyzed in this paper. Here we summarize the properties of the new protocols:

- Non-repudiation of the service by roamers.
- Incorrect bill cannot be charged on roamers.
  (The above two features are based on the assumption that service providers do not conspire.)
- A compromised session key does not compromise the contents of the following sessions, nor does it lead to the masquerading of either the subscriber or the service provider.
- Compromise of a network does not affect subscribers of others.
- Computation on the subscriber is simple, i.e., the battery of the mobile unit can last longer.
- Number of interactions among the subscriber and the networks are minimized. Together with previous property, it speeds up the verification and therefore reduces the total delay of validation.

REFERENCES

[1] A. Aziz and W. Diffie, "Privacy and Authentication for Wireless Local Area Network", IEEE Personal Communications, First Quarter, 1994, pp. 25-31.

[2] M. J. Beller, L. Cheng, Y. Yacobi, "Privacy and Authentication on a Portable Communication System", *IEEE J. on Selected Areas in Communications*, Vol. 11, No. 6, pp. 821-829, Aug. 1993.

[3] Dan Brown, "Security Planning for Personal Communications", *Proc. of 1st ACM Conference on Computer and Communications Security*, pp. 107-111, Nov. 1993.

[4] M. Burrows, M. Abai, and R. Needham, "A Logic of Authentication", ACM Transaction on Computer Systems, Vol. 8, No. 1, Feb. 1990, pp. 18-36.

[5] U. Carlsen, “Optimal Privacy and Authetication on a portable Communications System”, *Operating Systems Review*, June, 1994.
[6] Cellular Digital Packet data (CDPD) System Specification, Release 1.0, July 19, 1993.
[7] J.C. Cook, and R.L. Brewster, "Cryptographic Security Techniques for Digital Mobile Phones", *IEEE International Conference on Selected Topics in Wireless Communications*, pp. 425-428, 1992.
[8] S. Eckelman, "Minimizing Fraud", Telephone Engineering and Management, Vol. 94, No, 18, pp. 62-64, Sept. 1990.
[9] EIA/TIA-IS-54-B
[10] ETSI, ETS 300 175-7, October 1992.
[11] ETSI/TC Recommendation GSM 03.20, *Security Related Network Function*, version 3.3.2, Jan. 1991.
[12] H.Y. Lin and L. Harn, "Authentications in Wireless Communications," *Proc. of GLOBECOM '93*, pp. 550-554, Nov. 29-Dec. 2, 1993.
[13] R. Molva, D. Samfat, and G. Tsudik, “Authentication of Mobile Users”, *IEEE Network*, pp. 26-34, March/April, 1994.
[14] R. L. Rivest, A. Shamir, and L. Adleman, "A Method for Obtaining Digital Signatures and Public-key Cryptosystem", *Comm. of. ACM*, Vol. 21, No. 12, 1978, pp. 120-126.
[15] M. Shand and J. Vuillemin. Fast implementations of RSA cryptography. In Proceedings of the 11th IEEE Symposium on Computer Arithmetic, p.p. 252--259, IEEE Computer Society Press, Los Alamitos, CA, 1993.
[16] M. Spreitzer and M. Theimer, “Scalable, Secure, Mobile Computing with Location Information”, *Communications of the ACM*, Vol. 36, Iss. 7, p. 27, July, 1993.
[17] K. Vedder, “Security Aspects of Mobile Communication”, Computer Security and Industrial Cryptography - State of the Art and Evolution, East Course, Springer-Verlag, May, 1991, pp. 193-210.
[18] M. Walker, “Security in Mobile and Cordless Telecommunications”, Computer Systems and Software Engineering , Proceedings of CompEuro 1992, 1992, pp. 493-496.