A novel localised authentication scheme in IEEE 802.11 based Wireless Mesh Networks

Xiaodong Lin, Xinhua Ling, Haojin Zhu, Pin-Han Ho and Xuemin Sherman Shen*

Department of Electrical and Computer Engineering,
University of Waterloo, Canada
E-mail: xdlin@bbcr.uwaterloo.ca
E-mail: x2ling@bbcr.uwaterloo.ca
E-mail: h9zhu@bbcr.uwaterloo.ca
E-mail: pinhan@bbcr.uwaterloo.ca
E-mail: xshen@bbcr.uwaterloo.ca
*Corresponding author

Abstract: In the paper, we propose an efficient two-factor localised authentication scheme for inter-domain handover and roaming in IEEE 802.11 based service-oriented *wireless mesh networks* (WMNs). Some important aspects, such as resource-constraint *Mobile Stations* (MSs) and the ping-pong movement phenomenon when handover roaming across different hotspots occurs, are considered. An analytic model is developed to investigate the efficiency of the proposed scheme. Numerical results are given to demonstrate the superiority of the proposed scheme in terms of the resultant signalling overhead, power consumption, and authentication latency, compared with the legacy authentication schemes without losing the capability of preserving the system security.

Keywords: authentication; roaming; Wireless LANs; Wireless Mesh Networks; WMNs; hotspots.

Reference to this paper should be made as follows: Lin, X., Ling, X., Zhu, H., Ho, P-H. and Shen, X.S. (2008) 'A novel localised authentication scheme in IEEE 802.11 based Wireless Mesh Networks', *Int. J. Security and Networks*, Vol. 3, No. 2, pp.122–132.

Biographical notes: Xiaodong Lin is currently working toward his PhD Degree in the Department of Electrical and Computer Engineering at the University of Waterloo, Ontario, Canada, where he is a Research Assistant in the Broadband Communications Research (BBCR) Group. His research interest includes wireless network security, applied cryptography, and anomaly-based intrusion detection.

Xinhua Ling received the PhD Degree (2007) in Electrical and Computer Engineering at the University of Waterloo, Canada. His general research interests are in the areas of cellular, WLAN, WPAN, mesh and ad hoc networks and their internetworking, focusing on protocol design and performance analysis.

Haojin Zhu received the BEng Degree from Wuhan University, China, in 2002 and the MSc Degree from Shanghai Jiao Tong University, China, in 2005, all in Computer Science. He is currently pursing his PhD Degree in the Department of Electrical and Computer Engineering at the University of Waterloo, Ontario, Canada. His current research interests include wireless network security and applied cryptography.

Pin-Han Ho received his BSc (1993) and MSc Degree (1995) from the Electrical and Computer Engineering department at the National Taiwan University. He started his PhD study in 2000 at Queen's University, Kingston, Canada, focusing on optical communications systems, survivable networking, and QoS routing problems. He finished his PhD in 2002, and joined the Electrical and Computer Engineering department at the University of Waterloo, Ontario, Canada, as an Assistant Professor at the same year. He is the first author of more than 40 refereed technical papers and the co-author of a book on optical networking and survivability.

Xuemin Sherman Shen is a Professor of Electrical and Computer Engineering and University Research Chair at the University of Waterloo, Canada. He received his PhD in Electrical Engineering from Rutgers University, NJ, in 1990. His research interests include wireless/internet interworking, radio resource and mobility management, WLAN/WiMAX, UWB wireless communications, wireless ad hoc and sensor networks, wireless network security.

Copyright © 2008 Inderscience Enterprises Ltd.

1 Introduction

IEEE 802.11 (or Wi-Fi) based wireless hotspots, also known as public *Wireless LANs* (WLANs), have been on an upswing, and broadband wireless internet access has been readily available widely across metropolitan areas through wireless *Access Points* (APs), especially in heavily populated areas such as airports, restaurants, cafés, libraries, and hotels. Unlike the traditional Global System for Mobile (GSM Communications) or CDMA cellular networks, a typical Wi-Fi hotspot has a much smaller coverage and more restricted scalability, by which many different design requirements have been imposed in the efforts of interworking in such a heterogeneous network environment for achieving interoperability, security and cost-effectiveness without impairing the performance of throughput. This problem is getting more intractable when thousands of self-managed hotspots (Ohira et al., 2005) are operated by numerous different WISPs jointly provision broadband wireless access in a metropolitan area, and each hotspot may contain one or a number of APs which are physically accessed by the MUs.

In general, a MU subscribes to a WISP, called *home WISP*, and signs up an account in order to have access to the wireless internet services at the hotspots managed by the WISP. However, the MU cannot have access to a hotspot operated by any other WISP (called *foreign WISP*) unless the *foreign WISP* and the *home WISP* have a cooperative roaming agreement which stipulates how each other's users can have access to the hotspots managed by itself when a roaming event occurs. One of the most important issues associated with the WLAN roaming is the ability of providing a secure, light-weight, and cost-effective approach in authenticating the MUs' login requests, along with the consideration on the limitations of wireless communications environments, such as the limited computation power, memory, battery capacity of the MSs, and the ping-pong movement phenomena (Fang and Chlamtac, 1999).

Two authentication and roaming architectures in the interworking of WLANs have been widely employed (Laat et al., 2000; Zhang and Fang, 2006; 2007; Leu, 2006). One is by way of an *Authentication, Authorisation, and Accounting* (AAA) server (Laat et al., 2000) which coordinates the login requests from all the WLANs of the corresponding WISP. The foreign WISP responds to a MU's roaming request by sending the MU's authentication credentials (such as the MU's name, domain name, and password) to the AAA server of the MU's home network. The AAA server of the home network then authenticates the MU based on the received credentials and sends the authentication decision back to the foreign network (Anton et al., 2003). If successful, the MU will be granted the access right to the foreign network. The other roaming architecture is by way of a roaming broker (Leu et al., 2006), which is an authenticating agent trusted by all the WISPs in a certain region (such as in a metropolitan area), in order to exchange the roaming accounting information and deal with MUs' authentication requests. Unfortunately, the two authentication architectures are subject to a significant amount of signalling overhead among the roaming MU, the foreign network, the home network, and the AAA server or the roaming broker. Therefore, it is desired to have a light-weight authentication mechanism in order to reduce the roaming overhead. *Localised authentication* (Long et al., 2004; Baek et al., 2006) can meet the above design requirements, which is envisioned to serve as a key role in enabling the service-oriented wireless mesh networking applications in metropolitan areas. With localised authentication, an initial mutual authentication (also known as two-way authentication) between the roaming MU and the foreign network can be performed without any intervention of the MU's home network, which can significantly reduce the roaming and handover delay such that real-time services such as VoIP can be sufficiently supported in the presence of a high user handover frequency due to the small coverage of each hotspot.

Some studies have been reported to achieve localised authentication for both intra- and inter-network roaming across APs. An authentication protocol based on a public-key certificate structure was proposed for WLAN interworking by Long et al. (2004). The efficiency of the protocol comes from the fact that the signalling overhead between the AAA server and the WISPs can be avoided. Baek et al. (2006) proposed a localised AAA protocol to retain the mobility transparency from the AAA server in order to reduce the cost of the AAA procedure. In addition to providing mutual authentication, the authors demonstrated that the protocol can defend various threats such as replay attack, man-in-the-middle attack, and key exposure. Shi et al. (2007) realised localised authentication by using ticket, which is secrecy kept by MSs and APs after their first time authentication.

In this paper, we propose a localised authentication scheme for interworking of multiple WLANs in a metropolitan area operated by different WISPs. Our work is different from (Long et al., 2004; Baek et al., 2006; Shi et al., 2007) in the sense that an embedded two-factor authentication mechanism is implemented to determine whether a roaming MU is authentic without any intervention of the MU's home network as well as exchanging the authentication key in case the MU is authentic. A two-factor authentication mechanism is considered as a much stronger authentication method than the previously reported schemes since two independent authentication methods are adopted to ensure a MU to be a legitimate user, such as 'something you know' as one, and 'something you have' or 'something you are' as the other. In addition, the proposed localised authentication scheme has each WLAN cache the session key of each residing MU such that the impact due to the

ping-pong movement phenomena can be significantly mitigated. In particular, we consider the scenario where the MUs could roam between adjacent hotspots operated by common and different WISPs, which are differentiated in the cryptographic design of the authentication messages. An analytical model is developed to gain a deeper understanding on the performance of the proposed localised authentication scheme.

The remainder of this paper is organised as follows. In Section 2, we present a two-factor localised authentication scheme for WLAN roaming and interworking. In Sections 3 and 4, the security and efficiency of the proposed scheme are analysed and discussed, respectively. Finally, Section 5 concludes the paper.

2 Two-factor localised authentication scheme

In this section, a two-factor localised authentication scheme is proposed for the interworking of WLANs with multiple WISPs. The proposed localised authentication scheme based on Rabin cryptosystem aims to achieve efficient mutual authentication and session key exchange operations between MUs and APs.

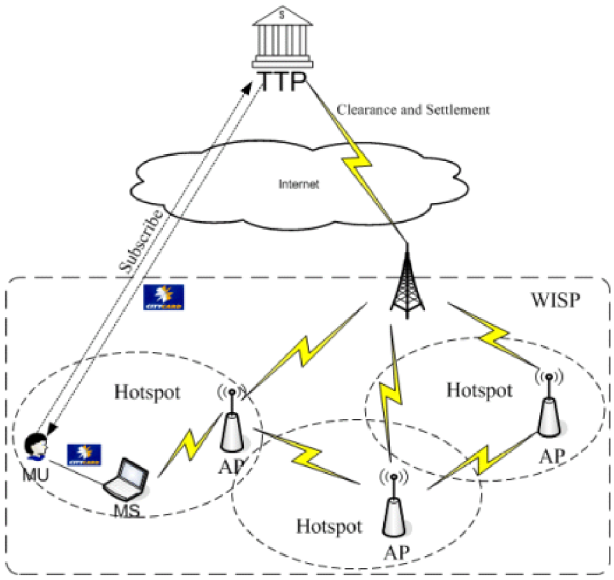
2.1 Rabin cryptosystem

Rabin cryptosystem is notably characterised by its asymmetric computational cost (Rabin, 1979), where the encryption (or signature verification) operation is extremely fast, while the decryption (or signature) operation is comparably slow and requires a large amount of computation effort (Gaubatz et al., 2005). This property makes Rabin cryptosystem rather suitable in a heterogeneous wireless network environment such as the interworking of WLANs, in which each AP of a hotspot is assumed to have unlimited power and sufficient computational capacity while the MSs have relatively limited resources. Thus, with the Rabin cryptosystem, an authenticated key agreement scheme can be devised such that the device heterogeneity in computational resources and battery capabilities are considered to properly allocate the cryptographic burden across the network domain. We refer to Rabin (1979) and Williams (1980) for a more comprehensive description of how Rabin's encryption and signature operations work.

2.2 Network architecture

For the considered network architecture, there are four types of network entities: the MUs, the *Trusted Third Party* (TTP), the WISPs, and the hotspots, while their relationship is shown in Figure 1. A WISP may operate multiple hotspots, which may or may not be adjacent to each other. Without loss of generality, we assume that each hotspot has a single AP, which makes the terms 'AP' and 'hotspot' interchangeable in the following context. The MUs can request for wireless internet access by subscribing to the TTP which has a mutual agreement with each WISP such that a subscribed MU can access to the hotspot operated by the corresponding WISP.

Figure 1 Network architecture



In addition to the role taken by the roaming broker, the TTP also serves as a trusted *Certificate Authority* (CA) server to issue certificates to both WISPs and MUs. The certificate issued to a MU or a WISP is a digital signature signed by the TTP on its public key as well as the linkage between the public key and the MU's or WISP's identity, respectively. Based on the authentication architecture, a *smart card* is issued by the TTP to each MU, which contains necessary authentication credentials for the MU and can serve as an *electronic pass* for the MU to roam across hotspots of different WISPs.

The proposed localised authentication scheme consists of the following three phases: the *initialisation* phase, the *login and mutual authentication* phase, and the *handoff* phase.

2.3 Initialisation phase

- *TTP initialisation*. TTP randomly chooses two primes, denoted as p_T and q_T, that satisfy $p_T, q_T \equiv 3 \pmod 4$, and a random number a_0 such that the Jacobi symbol $(a_0/n_T) = -1$ where $n_T = p_T q_T$. TTP then selects a randomly chosen large prime p and a generator g of the multiplicative group Z_p^*. Let $H(.)$ be a collision-resistant hash function. The quadruplet (a_0, n_T, p, g) together with the hash functions $H(.)$ are published as the public key, and the duplet (p_T, q_T) is kept as the private key.
- *WISP initialisation*. A WISP has to set up a mutual agreement with the TTP such that a legitimate MU can log in and have access to the wireless internet services provided by the WISP. In the proposed network architecture, a WISP has to perform the following operations:

- The WISP randomly chooses p_W, q_W that satisfy $p_w, q_w \equiv 3 \pmod 4$, and sends $n_W = p_W q_W$ to the TTP with its chosen identity ID_W while (p_W, q_W) is kept as its private key. The TTP checks the legitimacy of the identity and makes sure of the uniqueness of the WISP identity. If failed, the WISP has to choose another identity in order to proceed further.
- The TTP then computes two binary numbers c_1 and c_2, which will be used to sign the public key n_W of the WISP as well as the linkage between the public key n_W and the WISP's identity ID_W. c_1 can be determined by:

$$c_1 = \begin{cases} 0, & \text{if} \left(\dfrac{H(ID_W, n_W)}{n_T} \right) = 1 \\ 1, & \text{if} \left(\dfrac{H(ID_W, n_W)}{n_T} \right) = -1. \end{cases}$$

The TTP computes $t_0 = a_0^{c_1} \times H(ID_W, n_w)$, and derives c_2 such that:

$$c_2 = \begin{cases} 0, & \text{if} \left(\dfrac{t_0}{p_T} \right) = \left(\dfrac{t_0}{q_T} \right) = 1 \\ 1, & \text{if} \left(\dfrac{t_0}{p_T} \right) = \left(\dfrac{t_0}{q_T} \right) = -1. \end{cases}$$

Then, the TTP computes $r_0 = (-1)^{c_2} \cdot a_0^{c_1} \cdot H(ID_W, n_W) (\mathrm{mod}\, n_T)$, and derives s_0 such that $s_0^2 \equiv r_0 (\mathrm{mod}\, n_T)$.

- The TTP sends $(a_0, n_W, ID_W, s_0, c_1, c_2)$ to the WISP as its signature, which completes the WISP initialisation.

- *Hotspot initialisation*. A hotspot needs to be initiated by the corresponding WISP. First of all, the WISP chooses a_1 such that the Jacobi symbol $(a_1/n_W) = -1$ where $n_W = p_W q_W$. Then, the hotspot takes the following actions:
  - The hotspot randomly chooses p_A, q_A that satisfy $p_A, q_A \equiv 3 \pmod 4$, and sends $n_A = p_A q_A$ to its *home WISP* with its chosen *service set identifier*, $SSID_A$, while the duplet (p_A, q_A) is kept as the private key of the hotspot. The WISP checks the legitimacy of $SSID_A$ to ensure the uniqueness of the service set identifier of the hotspot. If not, the hotspot has to choose another service set identifier that is legitimate and unique in order to proceed further. The WISP is responsible for the service set identifier check on its managed hotspots.
  - The WISP computes c_3 and c_4 in order to create the signature on the public key n_A of the hotspot as well as the linkage between the public key n_A and the hotspot's service set identifier, $SSID_A$. c_3 is determined by:

$$c_3 = \begin{cases} 0, & \text{if} \left(\dfrac{H(SSID_A, n_A)}{n_W} \right) = 1 \\ 1, & \text{if} \left(\dfrac{H(SSID_A, n_A)}{n_W} \right) = -1. \end{cases}$$

Then the WISP computes $t_1 = a_1^{c_3} \times H(SSID_A, n_A)$, and derives c_4 such that

$$c_4 = \begin{cases} 0, & \text{if} \left(\dfrac{t_1}{p_W} \right) = \left(\dfrac{t_1}{q_W} \right) = 1 \\ 1, & \text{if} \left(\dfrac{t_1}{p_W} \right) = \left(\dfrac{t_1}{q_W} \right) = -1. \end{cases}$$

Finally, the WISP computes $r_1 = (-1)^{c_4} \cdot a_1^{c_3} \cdot H(SSID_A, n_A) (\mathrm{mod}\, n_W)$, and derives s_1 such that $s_1^2 \equiv r_1 (\mathrm{mod}\, n_W)$.

  - The WISP sends $(a_0, n_W, ID_W, s_0, c_1, c_2, a_1, n_A, SSID_A, s_1, c_3, c_4)$ to the hotspot, which completes the hotspot initialisation.

- *MU registration*. The roaming MU needs to subscribe to the TTP directly or through its home WISP in order to gain the wireless internet access. In the registration phase, the MU has to provide its billing information in order to continue the existing services. The registration process is detailed as follows:
  - The MU randomly chooses a password PW and computes $H(PW)$, and picks up a random identity ID_u. ID_u is a MU-controlled unique representation, called *virtual identity* (VID). The VID is only used for the purpose of billing. In this case, user privacy can be preserved by not disclosing the real identity of the MU. Then, the MU computes $P_u = g^{H(PW)} \bmod p$. Afterwards, the MU sends its credentials to the TTP for registration along with the two parameters P_u and ID_u through a secure channel (e.g., SSL). The user credentials submitted by the MU can be the billing address and payment method, etc. The identity should be unique for each MU.
  - The TTP specifies the expiry date for the MU, denoted by *expireDate*. Then, the TTP computes c_5 and c_6 as follows:

$$c_5 = \begin{cases} 0, & \text{if} \left(\dfrac{H(ID_u, P_u, \text{expireDate})}{n_T} \right) = 1 \\ 1, & \text{if} \left(\dfrac{H(ID_u, P_u, \text{expireDate})}{n_T} \right) = -1. \end{cases}$$

The TTP computes $t_2 = a_0^{c_5} \times H(ID_u, P_u, \text{expireDate})$ and derives c_6 such that

$$c_6 = \begin{cases} 0, & \text{if } \left(\frac{t_2}{p_T}\right) = \left(\frac{t_2}{q_T}\right) = 1 \\ 1, & \text{if } \left(\frac{t_2}{p_T}\right) = \left(\frac{t_2}{q_T}\right) = -1. \end{cases}$$

Then, the TTP computes $r_2 = (-1)^{c_6} \cdot a_0^{c_5} \cdot H(ID_u, P_u, \text{expireDate}) (\text{mod}\, n_T)$, and derives s_2 such that $s_2^2 \equiv r_2 (\text{mod}\, n_T)$.

- The TTP provides a smart card as an *electronic pass* to the MU through a secure approach, e.g., by registered mail, which is very similar to issuing a credit card. The smart card contains $(p, g, a_0, n_T, P_u, ID_u, \text{expireDate}, s_2, c_5, c_6)$

The proposed scheme employs a two-factor authentication mechanism to determine whether a roaming user is authentic. This is a necessary effort for security assurance improvement for localised authentication because it is the MU to provide confidential information for authentication while roaming across hotspots. Since the confidential information is at the user's side, leakage of the credentials may occur due to insufficient privacy protection or any malware such as keylogger, which results in intrinsically weakened security assurance. Thus, the second factor of authentication, which is realised by a smart card, can be addressed to compensate the weakness due to the adoption of localised authentication. This is based on the fact that if an attacker intends to impersonate a legitimate subscriber, the attacker has to not only know the authentication credentials of the legitimate subscriber, but also capture or replicate its smart card. It is well-known that the smart card industry has already made a significant progress in counteracting various attacks, such as physical attacks (Hendry, 1979). Obviously, by introducing an additional factor in the authentication process of MUs by way of the smart card device, the security of the proposed localised authentication scheme can be improved significantly.

However, it is very likely that a MU leaks its authentication credential and/or a smart card is lost or stolen. Thus, two appropriate actions have to be taken, which are shown as follows:

- MUs can change their passwords whenever they feel necessary by submitting their old password and newly chosen password to the TTP through a secure channel (e.g., SSL). The TTP signs on MUs' new credential, and sends back to the MUs. Upon receiving the updated parameters from the TTP, the smart card replaces the old parameters with the new ones. Thus, the password update completes.
- In the event of a loss or stolen of a smart card, the card holder should report the incident to the TTP. If the card is still valid, notification of a card revocation must be spread to all APs as rapidly as possible. However, a revocation will likely be a rare occurrence. Therefore, a simple card revocation process is built as follows: The TTP notifies all WISPs the incident. WISPs broadcast a card revocation message to all of their managed APs. Upon receiving the message, APs update their local *Card Revocation List* (CRL) with newly received revocated card information. For each AP, those expired cards are removed from the CRL periodically.

2.4 Login and mutual authentication phase

In this phase, a MU authenticates itself to an AP and the AP authenticates itself to the MU in such a way that both parties are assured of the others' legitimacy whenever the MU wishes to gain wireless internet access. Let AP have a service set identifier denoted as $SSID_A$. The access process is described as follows:

Step 1: Let each AP broadcast its public parameters, $(a_0, n_W, ID_W, s_0, c_1, c_2, a_1, n_A, SSID_A, s_1, c_3, c_4)$ periodically to all the associated MUs. In this case, the MU can easily ensure the security of the AP's public key n_A after validating the WISP's signature on it and the TTP's signature on the WISP's public key:

$$\begin{cases} S_0^2 = (-1)^{c_2} \cdot a_0^{c_1} \cdot H(ID_W, n_W) \bmod n_T \\ S_1^2 = (-1)^{c_4} \cdot a_1^{c_3} \cdot H(SSID_A, n_A) \bmod n_W. \end{cases}$$

If the validation fails, the MU aborts the login process since the MU could be subject to impersonation attack by the AP.

Step 2: The MU inserts its smart card into the smart card reader of the mobile device and enters its password, *PW*. The smart card will perform the following operations: 1) calculate $H(PW)$, and 2) generate two random numbers x and r_1, where x is a k-bit number taken as the MU's key contribution. The smart card randomly chooses a key k_{ma}, and encrypts the key k_{ma} by using the following relation:

$$l_1 = k_{ma}^2 \bmod n_A$$

Then, the smart card encrypts its identity ID_u, x, r_1, P_u, *expireDate*, s_2, c_5, c_6 and T by using the following relation:

$$l_2 = E_{k_{ma}}(ID_u, x, r_1, P_u, expireDate, s_2, c_5, c_6, T)$$

where $E_k(m)$ means encryption of message m by using any implicit secure symmetric encryption algorithm under the key of k, T is the current date and time of smart card reader. Then, the MU sends a login request l_1, l_2 to the AP for the authentication process.

Step 3: With a login request, the AP decrypts l_1 with its private key (p_A, q_A) to obtain the key k_{ma}, and then obtains $(ID_u, x, r_1, P_u, expireDate, s_2, c_5, c_6, T)$ by decoding l_2 using newly deciphered key k_{ma}. Then, the AP takes the following actions:

- Checks whether the MU's account has expired according to *expireDate*.
- Checks against its list of revoked cards, and if found, it stops and rejects the login request. Otherwise, continue.

- Checks the format of ID_u, and will reject this login request if it is not correct.
- Checks if the timestamp T is reasonable, and if so, continue. Otherwise, it stops and rejects the login request.
- Verifies whether $s_2 = (-1)^{c_6} \cdot a_0^{c_5} \cdot H(ID_u, P_u, \exp ireDate) \bmod n_T$ holds. If it holds, the AP accepts the login request. Otherwise, the login request will be rejected.

The AP randomly chooses a k-bit number y as the AP's key contribution, and computes $d = E_x(SSID_A \| y)$. Then, the AP sends d back to the MU with a random challenge r. Also, the AP calculates a session key by using the following: $K = H(ID_U \| SSID_A \| x \| y)$..

Step 4: After the MU receives (d, r), the MU decrypts d by using symmetric-key decryption: $E_x^{-1}(d) = SSID_A \| y$ in order to verify $SSID_A$, where a session key is obtained by the following computation: $K = H(ID_U \| SSID_A \| x \| y)$. Then, the MU randomly chooses a number k such that $0 < k < p-1$ and gcd(k, $p-1$) = 1, and computes $v = g^k \pmod p$. Note that the above calculation can be done in advance, and the result can be stored at the MU side in order to speed up the login process. With k and v, the MU computes $w = (H(r) - H(PW)v)k^{-1} \pmod{p-1}$. If $w = 0$, the MU starts over again by choosing a different k. Afterward, the MU computes $c_u = E_K(w, v)$ and sends c_u to the AP.

Step 5: Upon receiving c_u, the AP can correctly recover (w, v) by using newly generated session key K. Then, the AP checks if $g^{H(r)} \equiv P_u^v v^w$. If it holds, the AP accepts the MU as a legitimate user, and accepts the MU to login the system. The AP also establishes a session key K with the MU. An entry of ($sessionID, ID_U, K, r, w, P_u, v, T_u, \exp ireDate, Lifetime$), called *user information record* (*UIR*), is kept in the AP's local database, where T_u records the usage of the MU's wireless internet service at the AP, and *Lifetime* serves as a timer controlling how long the entry is active. If the timer hits 0, the entry is expired. Afterwards, the confidentiality and integrity of communication between the MU and the AP is protected by the session key K. Also, the MU keeps ($sessionID, SSID_A, K, Lifetime$) in its local cache table corresponding to the session key K, which aims to reduce the impact of ping-pong movement effect.

2.5 Handoff phase

While a communication session is on, the MU may move from the AP to an adjacent hotspot, denoted as AP_2, which is operated by the same or different WISPs. The handoff procedure is performed between the MU and AP_2 to ensure that the on-going service is not interrupted while preventing any unauthorised access. In order to further improve the efficiency of the proposed localised authentication scheme, the session key is cached in the current network domain. This is considered to be able to effectively mitigate the impact by any possible ping-pong movement phenomenon which is inevitable in any motion prediction and handoff algorithm. Whenever a MU requests a handoff into an AP which has a non-expired shared session key with the MU, a user authenticated key agreement protocol with secret-key cryptography will take place instead of a full authentication procedure introduced above by using asymmetric key encryption, which are shown as follows:

Step 1: Let AP_2 have already broadcast its public parameters $(a_0, n_W, ID_W, s_0, c_1, c_2, a_1, n_A, SSID, s_1, c_3, c_4)$ to all the associated MUs. In this case, the MU can easily ensure the security of AP_2's public key n_A after validating WISP's signature on it and the TTP's signature on WISP's public key.

$$\begin{cases} S_0^{\,2} = (-1)^{c_2} \cdot a_0^{\,c_1} \cdot H(ID_W, n_W) \bmod n_T \\ S_1^{\,2} = (-1)^{c_4} \cdot a_1^{\,c_3} \cdot H(SSID, n_A) \bmod n_W . \end{cases}$$

If the validation fails, the MU will abort the login process since the MU could experience the attack of impersonation by the AP.

Step 2: Based on the SSID of AP_2, the MU looks for the session ID and session key of AP_2 in its cache table. If there does not exist a non-expired session key, a regular authenticated key agreement and login process will be performed, but it is unnecessary for the MU to enter its password during the handoff phase because the smart card has cached the hash value of the MU's password as long as the smart card is still inserted. In case that a valid session key can be found, the MU chooses a random nonce n_1 and a random k-bit number x' as the MU's new session key contribution. Then, the MU computes $y_1 = MAC_K(sessionID \| n_1)$ and $e_1 = E_K(x')$, where K is the cached old session key between the MU and AP_2. Afterwards, the MU sends ($sessionID$, n_1, y_1, e_1) to the AP_2.

Step 3: AP_2 checks if the session with *sessionID* has expired. If so, it requests the MU to perform a regular authenticated key agreement and login process. Otherwise, it verifies $MAC_K(sessionID \| n_1)$ by using the cached session key. If succeeds, the AP_2 continues and recovers x'. Otherwise, it is aborted. Then, AP_2 chooses a random nonce n_2 and a random k-bit number y' as AP_2's new session key contribution, and computes $y_2 = MAC_K(SSID \| n_1 \| n_2)$ and $e_2 = E_K(y')$. Afterwards, AP_2 sends (n_2, y_2, e_2) to the MU.

Step 4: Upon receiving the triplet (n_2, y_2, e_2), the MU verifies $MAC_K(SSID \| n_1 \| n_2)$. If succeeds, the MU continues and recovers y'. Otherwise, it is aborted. Afterward, both the MU and AP_2 can calculate the new session key $K' = H(ID_U \| SSID \| x' \| y')$, and the MU's user information record at AP_2 will be updated accordingly with the newly generated session K'. It can be observed

that the simplified version of authenticated key agreement protocol only uses symmetric-key encryption and keyed-hash message authentication codes which are very fast operations compared with asymmetric-key encryption.

3 Security analysis

In this section, the security of the proposed scheme is analysed.

1 *Prevention of replay attacks*. With a replay attack, an adversary replays the intercepted login message in order to impersonate as a legitimate user. Obviously, it cannot work in the proposed scheme because of the time interval check in Step 3. If timestamp T included in a login request message is not reasonable, the AP will simply reject the login request.

2 *Prevention of impersonation attack*. In the application scenario considered in the study, an impersonation attack can be either on a legitimate MU or on an AP. The mutual authentication mechanism has been used in the proposed scheme to prevent the impersonation attack in both situations. Firstly, an adversary cannot impersonate a legal login request even it has intercepted any previous login message. To forge a valid login, the adversary needs to have both the valid smart card and the corresponding password. Note that the password is difficult to derive by using the knowledge of public parameter P_u, where $P_u = g^{H(PW)} \bmod p$, and PW is the password. The security guarantee relies on the difficulty of computing discrete logarithms over finite fields, which is considered to be difficult. Furthermore, the one-way hash function is used to hide the real user password, which is computationally infeasible to reveal the user password by only knowing $H(PW)$, which is the hash value of the password. Thus, the adversary cannot forge a valid login and impersonate a legitimate MU. Secondly, the proposed localised authentication scheme prevents an impersonation attack upon the AP. A malicious attacker could broadcast bogus beacons to attract the legitimate users in its radiation range, which could possibly defraud the legitimate users for their authentication information. In the proposed scheme, a highly efficient mutual authentication is devised to resist this attack, where the MU sends a challenge $l_1 = k_{ma}{}^2 \bmod n_A$ and $l_2 = E_{k_{ma}}(ID_u, x, r_1, P_u, expireDate, s_2, c_5, c_6, T)$ to the AP after verifying the public key of the AP. Only a real AP with the knowledge of factorisation of n_A can compute $ID_u, x, r_1, P_u, expireDate, s_1, c_5, c_6, T$ and respond correctly with $d = E_x(SSID_A \| y)$ by using the secret key x, which is contained inside the challenge l_2 from the MU. The security guarantee relies on the difficulty of integer factorisation, which is considered to be difficult as well.

3 *Prevention of password guessing attack*. It is difficult for any adversary to derive the user password without knowing the private key of the AP. Moreover, an adversary cannot perform an offline password guessing attack because a random number r_1 is introduced in $l_2 = E_{k_{ma}}(ID_u, x, r_1, P_u, expireDate, s_2, c_5, c_6, T)$.

4 *Multiple factor authentication*. In reality, it may happen that the MU leaks its password. However, even if the adversary knows the MU's password, it still can not forge a user login. To create a forged login, the adversary needs the information stored in the MU's smart card, which is hard to get. The smart card provides an extra layer of protection in the authentication procedure. In addition, once the MU comes to know that its password is leaked, the MU can easily invoke the password change protocol to change its password. Furthermore, in practice, the MUs can increase the level of security by changing their passwords frequently.

5 *Privacy of communication between the MU and AP*. The communication between the MU and the AP is protected by a secret session key K such that an adversary cannot know the content without the knowledge of the key K. To obtain K, the adversary needs to know not only the MU's key contribution x but also the AP's key contribution y. For the MU's key contribution x, only a party with the knowledge of factorisation of n_A can decrypt $l_2 = E_{k_{ma}}(ID_u, x, r_1, P_u, expireDate, s_2, c_5, c_6, T)$, which is an integer factorisation problem. For the AP's key contribution y, only a party with the knowledge of the MU's key contribution x can decrypt $d = E_x(SSID_A \| y)$, which is still equivalent to an integer factorisation problem.

In summary, the proposed scheme can resist the reply attack, impersonation attack, password guessing attack, and protect privacy of communication between the MU and AP. It also achieves multiple-factor authentication for much better assurance of security.

4 Performance evalution

We perform a series of experiments through qualitative discussions and computer simulation to evaluate the proposed authentication scheme, particularly under the impacts of the authentication latency and additional energy consumption caused by the handoff process. The considered network model used in the experiments is based on the proposed network architecture shown in Figure 1. Some WLANs have overlapping geographic areas, while others are completely separated. The MUs roam across the WLANs in the network, and handoff occurs when the MU roams from a currently serving WLAN to another. Note that in a conventional cellular system, handoff is the mechanism

of transferring an ongoing call from one serving cell to another or from the serving channel to another within the same cell. In the network considered here, on the other hand, handoff refers to transferring an ongoing connection from one WLAN to another. Furthermore, in the experiments, the resources at the MU and AP sides are taken asymmetric. At the MU side, the used mobile device has limited power and computational capacity. On the other hand, an AP is physically stationary with unlimited power and sufficient computational capacity.

4.1 Authentication latency

In the proposed scheme, fast authentication can be achieved based on the following two facts: firstly, the delay can be largely reduced due to the adoption of a localised authentication strategy; secondly, the proposed scheme is based on the ElGamal signature scheme (ElGamal, 1985) and Rabin cryptosystem (Rabin, 1979). The ElGamal signature scheme has been well-known for its capability in the pre-computation of some intermediate values in the signature signing process in an offline manner. The ElGamal signature generation operation can thus become very fast by going through only a single large modular multiplication. In the proposed scheme, the mobile device only needs to implement the ElGamal signature generation operation.

The Rabin cryptosystem can be executed in an extremely high speed at one side due to its asymmetric computational cost, where the encryption (or signature verification) operation is extremely fast, while the decryption (or signature generation) operation is comparably slow and requires a large amount of computation effort (Rabin, 1979). In the proposed scheme, only two signature verifications and one encryption operations of Rabin cryptosystem are performed by the mobile device, which can be executed in an extremely short time.

Table 1 shows the execution time of encryption and verification operations of Rabin cryptosystem, and the signature generation and verification operations of ElGamal signature scheme.

By observing Table 1, the computing time at the MU side is negligible. In addition, by way of caching the session keys during the roaming MU's handoff phase, the delay can be further reduced since only symmetric-key encryption and keyed-hash message authentication code operations are required in each authentication process, which is very fast and can be easily implemented.

4.2 Energy cost

Since the AP is assumed to have no power limitation, we only take the MU's energy consumption into consideration. We assume that the symmetric-key encryption algorithm, hash function, and keyed-hash message authentication code algorithms adopted in the system are AES, MD5 and HMAC, respectively. Table 2 shows the energy consumption of various cryptographic operations involved in the proposed scheme.

Table 1 Latency (in millisecond) of cryptographic operations of Rabin cryptosystem and ElGamal signature scheme

	Rabin-1024		*ElGamal-1024*
Encryption	0.015	Signature with no pre-computation	2.26
		Signature with pre-computation	0.46
Verification	0.015	Verification	2.74

*We evaluate the computational costs of cryptographic operations of Rabin cryptosystem and ElGamal signature scheme on an Intel Pentium 4 3.0 GHz machine with 1 GB RAM running Fedora Core 4 based on cryptographic library MIRACL (http://www.shamus.ie/).

Table 2 Energy cost of cryptographic operations (in mJ)

	Rabin-1024	*ElGamal signature scheme-1024*	*MD5*	*AES*	*HMAC*
Encryption cost	7.98	N/A	0.59 uJ/Byte	1.21 uJ/Byte	1.16 uJ/Byte
Ciphertext overhead (bits)	1024	N/A	128	N/A	128
Verification cost	7.98	338.02	N/A	N/A	N/A
Signature overhead (bits)	1024	2048	N/A	N/A	N/A
Signature generation cost	1596	313.60	N/A	N/A	N/A

*The cost of receiving one byte is 28.6 μJ, and the cost of transmitting a byte is 59.2 μJ (Potlapally et al., 2006).

Let E denote the total energy cost of one authentication procedure, which includes the following three components:

- various cryptographic operations
- the cost of transmission
- receiving exchanging messages between the MU and the AP.

There are two types of authentication processes in the proposed scheme. One is the regular authentication process, which happens during the MU's login phase or handoff phase without non-expired cached session key with the newly associated AP. The other is the caching session key authentication process, which occurs during the MU's handoff phase with non-expired cached session key with the newly associated AP.

Based on Table 2, the energy consumed in the two types of authentication is estimated to be 367.48 mJ and 31.24 mJ, respectively. It is observed that the energy consumption in an authentication procedure with the caching session key mechanism can be significantly reduced.

To further investigate the efficiency of the proposed localised authentication scheme, we develop an analytic model by considering some important issues specific to the wireless environments, especially on the ping-pong movement phenomena when the MUs roam across different APs.

4.2.1 Analytic model

Usually, the statistical information of *call holding time* T_c and *cell residence time* T_d are needed in analysing the performance of handoff schemes. Given the mean or distribution of T_c and T_d, the average number of handoffs that a call will experience in its life time can be obtained. Together with other information (e.g., the new call arrival rate, user mobility pattern, and the cell/WLAN capacity), important performance metrics such as new call blocking probability, handoff call dropping probability and call forced termination probability can also be obtained through classical queueing analysis (see Fang and Chlamtac, 1999; Xu et al., 2004 and the references therein).

For practical network planning, when real system measurements are not available, one can choose these parameters based on the average speed of MUs, their movement patterns, and the coverage area of WLANs. Once the network is deployed and traffic measurements are available, these parameters can be refined based on statistical analysis of such measurements. For example, one can collect data regarding the movement and call holding time of MUs, and obtain mean call holding time $1/\mu$ and mean cell residence time $1/\eta$ based on the statistical analysis of such data. Here the location of each MU, the time-of-the-day of characteristic of each WLAN, and any other factor that relates to traffic patterns can be taken into consideration. Such an approach is already used in traffic engineering and dimensioning of communication networks (Naghshinel and Schwartz, 1996). However, the details of these approaches are beyond the scope of this paper.

For the purpose of discussing the authentication cost, we assume that the call duration in WLANs is exponentially distributed with the mean ($1/\mu$) seconds, as commonly adopted in the study of handoff in cellular systems (Naghshinel and Schwartz, 1996; Xu et al., 2004). The time that a call spends in a WLAN prior to handing off to another, i.e., the dwell time, is assumed to be generally distributed with probability density function $f_{T_d}(x)$. Based on the above assumptions, we obtain the probability of having at least one more handoff for an ongoing call P_h:

$$\begin{aligned} P_h &= \Pr\{T_d \le T_c\} \\ &= \int_0^\infty f_{T_d}(x) \int_x^\infty f_{T_c}(y)\, \mathrm{d}y \mathrm{d}x \\ &= \int_0^\infty f_{T_d}(x) \int_x^\infty \mu \mathrm{e}^{-\mu y}\, \mathrm{d}y \mathrm{d}x \\ &= \int_0^\infty f_{T_d}(x) \mathrm{e}^{-\mu x}\, \mathrm{d}x \end{aligned} \tag{1}$$

where $f_{T_d}(x)$ is the probability density function of the call duration time. Note that P_h is constant regardless of the ongoing call's handoff history, due to the memorylessness of the exponentially distributed call duration.[1] Thus, the number N_h of handoffs that an ongoing call will experience during its call life time follows the geometric distribution with parameter P_h, i.e.,

$$\Pr\{N_h = k\} = (1 - P_h) P_h^k \qquad k = 0, 1, \ldots \tag{2}$$

4.2.2 Ping-pong effect

Handing off back and forth between two APs in a relatively short period of time is referred to as a *ping-pong effect*. This is usually due to the rapid fluctuation of the received signal strengths from the two APs, and usually occurs if the handoff scheme initiates a handoff request when the received signal strength of a new AP is stronger than that of the current AP (Thajchayapong and Peha, 2006). The ping-pong effect causes a problem of inducing unnecessary overhead in terms of signalling, authentication, and database update frequency, etc. In some situations, it may result in network instability and disruption (Naghshinel and Schwartz, 1996).

Some handoff schemes aiming to mitigate the ping-pong effect have been proposed in the literature. Basically, a hysteresis margin between the *Received Signal Strengths* (RSS) of the current and candidate BSs and timers are used (Pollini, 2006). One class of such schemes addresses that the RSS of the candidate foreign AP needs to be stronger than the RSS of the current AP with a hysteresis margin δ (dB) before a handoff request is initiated. Further improvements have been made by confirming that the link quality of the candidate AP is better. For many cases these mechanisms can be effective in reducing ping-pong effect, but cannot completely eliminate it due to different user movement patterns and environmental factors such as the layout of buildings.

From a quantitative point of view, we define that a ping-pong effect occurs if two consecutive handoffs take place within a relatively short period of time T_{pp}, called *ping-pong threshold*, which is described as follows:

γ= Pr{a handoff occurs within T_{pp} since the last handoff}

where γ is denoted as the ping-pong effect ratio. Then, among the handoff calls, approximately γ of them will be ping-pong handoff calls.

4.2.3 Numerical results

In the following, the impact of ping-pong effect on energy cost of the proposed localised authentication scheme is investigated, mainly based on the field measurements of *cell dwell time* in WLANs reported in Thajchayapong and Peha (2006). It is observed that the approximated distribution of dwell time is:

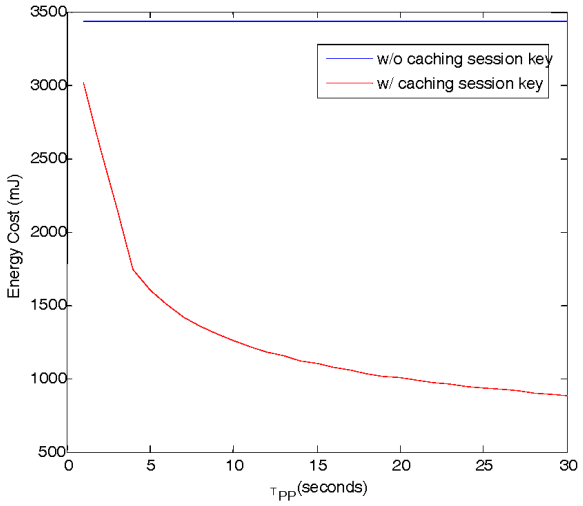
$$f_{T_d}(t) = \begin{cases} 0.135, & 0 \le t < 4 \\ 0.374 t^{-1.44}, & t \ge 4 \end{cases}. \tag{3}$$

Given μ, P_h can be obtained numerically by substituting equation (3) into equation (1). Meanwhile, γ can be obtained as

$$\gamma = \int_0^{T_{pp}} f_{T_d}(t)\,\mathrm{d}t. \tag{4}$$

The cost saving due to the adoption of session key caching in the proposed localised authentication scheme considering the ping-pong effect is shown in Figure 2. When the caching session key mechanism is applied, the energy cost can be reduced significantly when the ping-pong effect increases. In addition, the energy cost decreases less significantly after *ping-pong threshold* T_{pp} reaches around 4 s, especially, the saving of energy cost is kept steady after T_{pp} is greater than 30 s. In reality, the caching session key mechanism also introduces possible session key exposure, which could result in service theft and abuse. The longer a session key is cached, the higher risk MU could face key exposure. Therefore, we should increase the life time of a session key as long as we want, where a tradeoff between security requirement and efficiency must be initiated. Based on the experiment results in the study, we can limit the lifetime of a cached session key to around 5 s, which is the time point such that the rapid decrease of energy consumption by increasing T_{PP} just ends. However, except there comes up with a general cost function for the risk of key exposure, it will be hard for any optimisation task that can be performed for achieving the best compromise. This will be left as our future research.

Figure 2 Ping-pong effect ratio vs. energy cost



5 Conclusions

In this paper, we have introduced an efficient two-factor localised authentication scheme for IEEE 802.11 based WMN in metropolitan areas. The proposed scheme takes advantages of Rabin cryptosystem and ElGamal signature scheme, and is considered as a stronger authentication scheme than all the conventional authentication schemes, in order to compensate the increased impact caused by more heavily authorised APs. The proposed scheme has been developed by considering the constraints in the wireless communications environment such as limited computation power, memory and battery capacity of MSs, and ping-pong movement problem when roaming across WLANs. We have demonstrated that the proposed authentication process can not only achieve negligible roaming/handoff latency, but also significantly reduce the energy consumption at the MU side, which can serve as a key enabling technology for the future metropolitan-area WMNs.

References

Anton, B., Bullock, B. and Short, J. (2003) *Best Current Practices for Wireless Internet Service Provider (WISP) Roaming*, Wi-Fi Alliance, February.

Baek, S., Pack, S., Kwon, T. and Choi, Y. (2006) 'A localized authentication, authorization, and accounting (AAA) protocol for mobile hotspots', *Proceedings of the Third Annual Conference on Wireless on Demand Network Systems and Services*, Les Ménuires, France, January, pp.144–153.

ElGamal, T. (1985) 'A public key cryptosystem and a signature scheme based on discrete logarithms', *IEEE Transactions on Information Theory*, Vol. 31, No. 4, July, pp.469–472.

Fang, Y. and Chlamtac, I. (1999) 'Teletraffic analysis and mobility modeling of PCS networks', *IEEE Transactions on Communications*, Vol. 47, No. 7, July, pp.1062–1072.

Gaubatz, G., Kaps, J.P. and Sunar, B. (2005) 'Public key cryptography in sensor networks – revisited', *ESAS*, LNCS 3313, pp.2–18.

Hendry, M. (1997) *Smart Card Security and Applications*, Artech House Publishers, Norwood, MA, USA.

Laat, c.d., Gross, G., Gommans, L., Vollbrecht, J. and Spence, D. (2000) 'Generic AAA architecture', *RFC 2903*, Internet Engineering Task Force (IETF).

Leu, J., Lai, R., Lin, H. and Shih, W. (2006) 'Running Cellular/PWLAN services: practical considerations for Cellular/PWLAN architecture supporting interoperator roaming', *IEEE Communications Magazine*, Vol. 44, No. 2, July, pp.73–84.

Long, M., Wu, C-H. and Irwin, J.D. (2004) 'Localized authentication for wireless LAN Internetwork roaming', *Proceedings of IEEE Wireless Communications and Networking Conference*, Atlanta, GA USA, March, pp.264–267.

Naghshinel, M. and Schwartz, M. (1996) 'Distributed call admission control in mobile/wireless networks', *IEEE Journal on Selected Areas in Communications*, Vol. 14, No. 4, May, pp.711–717.

Ohira, K., Huang, Y., Okabe, Y., Fujikawa, K. and Nakamura, M. (2005) 'Security analysis on public wireless Internet service models', *Proceedings of the 3rd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots*, Cologne, Germany, September, pp.107–110.

Pollini, G.P. (1996) 'Trends in handover design', *IEEE Communications Magazine*, Vol. 34, No. 3, January, pp.82–90.

Potlapally, N.R., Ravi, S., Raghunathan, A. and Jha, N.K. (2006) 'A study of the energy consumption characteristics of cryptographic algorithms and security protocols', *IEEE Transactions on Mobile Computing*, Vol. 5, No. 2, February, pp.128–143.

Rabin, M.O. (1979) *Digitized Signatures and Public-key Functions as Intractable as Factorization*, Technical Report, Massachusetts Institute of Technology, Cambridge, MA, USA.

Shi, M., Rutagemwa, H., Shen, X.S., Mark, J.W. and Saleh, A. (2007) 'A service agent based roaming architecture for WLAN/Cellular integrated networks', *IEEE Transactions on Vehicular Technology*, Vol. 56, No. 4, July, pp.3168–3181.

Thajchayapong, S. and Peha, J.M. (2006) 'Mobility pattern in microcellular wireless networks', *IEEE Transactions on Mobile Computing*, Vol. 5, No. 1, January, pp.52–63.

Williams, H.C. (1980) 'A modification of the RSA public-key encryption procedure', *IEEE Transactions on Information Theory*, Vol. 26, No. 6, November, pp.726–729.

Xu, Y., Ding, Q. and Ko, C.C. (2004) 'Impact of handoff protection strategies on cellular mobile system capacity', *IEEE Transactions on Wireless Communications*, Vol. 3, No. 4, July, pp.1076–1087.

Zhang, Y. and Fang, Y. (2006) 'ARSA: an attack-resilient security architecture for multi-hop wireless mesh networks', *IEEE Journal on Selected Areas in Communications*, Vol. 24, No. 10, October, pp.1916–1928.

Zhang, Y. and Fang, Y. (2007) 'A secure authentication and billing architecture for wireless mesh networks', *ACM Wireless Networks*, Vol. 13, No. 5, October, pp.663–678.

Note

[1]If the call duration follows a distribution other than exponential, P_h will be a function of the number of handoffs experienced by the call.

Websites

GSM Association, WLAN roaming guidelines, at http://www.gsmworld.com/documents/wlan/ir61.pdf

Multiprecision Integer and Rational Arithmetic C/C++ Library (MIRACL), http://www.shamus.ie/