Formal Analysis of the Handover Schemes in Mobile WiMAX Networks

Ahmed M. Taha[1], Amr T. Abdel-Hamid[1], and Sofiène Tahar[2]
[1]Faculty of Information Engineering and Technology
German University in Cairo (GUC), Cairo, Egypt
{ ahmed.mohamedtaha, amr.talaat}@guc.edu.eg
[2] Department of Electrical and Computer Engineering
Concordia University, Montreal, Quebec, Canada
tahar@ece.concordia.ca

***Abstract*— An overview of the EAP-based handover procedures of the IEEE 802.16e standard is introduced and their security vulnerabilities are analyzed. Possible solutions for secure handover in IEEE 802.16e networks that guarantee a backward and forward secrecy are described and formally verified using Scyther, a specialized model checker for security protocols. These solutions showed a few drawbacks in the verified procedure and some modifications are proposed for a more secure and efficient handover protocol.**

***Index Terms*—Formal Verification, Handover, Mobile WiMAX, Pre-authentication Scheme, Security Protocol.**

I. INTRODUCTION

The IEEE 802.16 standard of the Worldwide Interoperability for Microwave Access (WiMAX), aims to provide broadband wireless access for Metropolitan Area Networks (MAN) and offers all packet-switched services for fixed, nomadic, portable, and mobile accesses. The first specification, IEEE 802.16-2004 [1], also called Fixed WiMAX, was ratified by IEEE in 2004 and was extended by the development of IEEE 802.16e [2], also called Mobile WiMAX, which supports mobility so mobile stations can handover between base stations while communicating, as well as supporting some other functions including multicast. In Mobile WiMAX, there are three possible approaches specified to implement handover (HO). Hard handover (HHO), Macro Diversity Handover (MDHO), and Fast Base Stations Switching (FBSS).

IEEE 802.16e supports two different mechanisms for authentication: the mobile station (MS) and the base station (BS) may use RSA-based authentication or Extensible Authentication Protocol (EAP)-based authentication. EAP-based authentication uses a backend infrastructure, such as the AAA (Authentication, Authorization, and Accounting) architecture. Due to the flexibility and ability to interact with AAA infrastructures, it is very likely that EAP will become the de facto authentication method for 802.16e access control [3].

Mobile-service users expect the handover procedure to be completed as quickly as possible, but on the other hand, authentication is a time-consuming operation, which is an essential component of the handover procedure. Therefore it is important to reduce the computation for encryption or decryption. On the other hand, EAP lacks the ability to support MS mobility, as a full EAP authentication latency requires about 1000 ms [4]. However, supporting voice and multimedia with mobility implies that the total handoff latency must be small. The recommended maximum handoff latency for voice over IP (VoIP) applications is 50 ms [5], and for streaming video/audio applications is 150 ms [6]. It is difficult to address the problem of low cost and quick handover. 802.16e security sub-layer does not offer a complete and effective solution.

In this paper, we provide an informal description of handover procedures in IEEE 802.16e standard and their security flaws. Afterwards, we discuss the potential solutions proposed for secure handover in mobile WiMAX networks and pre-authentication scheme proposed by Hur *et al* [7]. The pre-authentication scheme is analyzed informally and guarantees a backward and forward secrecy. We perform formal analysis and verification on handover procedures in IEEE 802.16e standard using the Scyther tool [8], in order to extract the main security flaws and threats that might exist in such procedures. Finally, we propose some modifications on the mechanisms verified to ensure the security of different handover procedures.

This paper is organized as follows. Section II introduces the IEEE 802.16e PKMv2 key management and handover mechanisms. In Section III, we describe the secure pre-authentication schemes for mobile WiMAX networks. In Section IV, we formally analyze the performance and security of the PKMv2-based handover schemes of IEEE 802.16e, as well as the pre-authentication scheme using Scyther, we also perform some modifications for a more efficient protocol. Finally in Section V, we conclude the paper.

II. INFORMAL ANALYSIS OF IEEE 802.16E PKMV2 HANDOVER

Privacy in IEEE 802.16e specification has two component protocols: An encapsulation protocol for encrypting packet data across the fixed broadband wireless access (BWA) network and a privacy key management (PKM) protocol providing the secure distribution of keying data from BS to MS. There are two versions of privacy key management

978-1-4244-3474-9/09/$25.00 ©2009 IEEE

protocols supported in IEEE 802.16e: PKMv1 and PKMv2 [2].

The PKMv1 is vulnerable to man-in-the-middle attack, and simple message replay attack, which will cause denial of service as well as the necessity of mutual authentication. These were the main reasons that led to the evolution of PKMv2 that caters the shortcomings of the first version. Thus, the security of the IEEE 802.16e is introduced in term of the PKMv2 in this paper.

A. PKMv2 in IEEE 802.16e

There are two authentication schemes in 802.16e, one based on RSA and the other based on EAP. In this paper, we focus on the EAP-based authorization, especially single EAP mode. EAP-based authentication uses a backend infrastructure as authentication server (AS), such as the AAA architecture.

The flow of message exchange and key derivation and delivery procedure using the EAP-based authentication mode are described in Fig. 1. At the initial entry, The MS authenticates to an AS via an authenticator. The BS in 802.16 networks serves as the authenticator. EAP authentication follows the steps listed in [9].

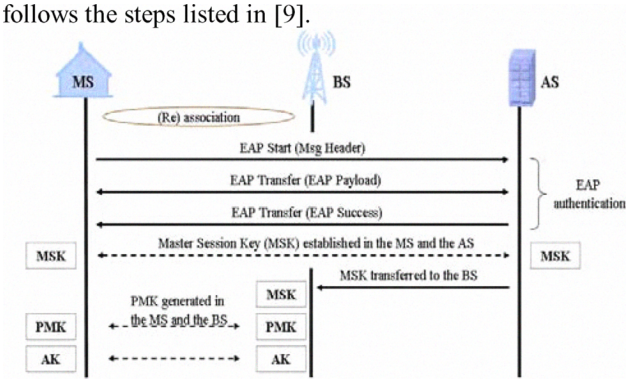

Fig. 1. PKMv1 message exchange and key derivation

At the end of the protocol run, the AS and the MS have the 512 bits master session key (MSK). The AS delivers the MSK to the BS after the EAP exchange is complete. Then the MS and the BS derive a pairwise master key (PMK) by truncating the MSK to 160 bits, and derive an authorization key (AK) from the PMK. PMK is used with the MS's MAC address (MSID) and the base station identifier (BSID) to generate a 160-bit AK, the generation process is shown as follows:

PMK = *Truncate* (MSK, 160),
AK = *Dot16KDF* (PMK, MSID / BSID / "AK", 160). (1)

The Dot16KDF algorithm is a counter mode encryption (CTR) construction that may be used to derive an arbitrary amount of secret key from source keying material [10].

AK is then used to generate the key encryption key (KEK) and hash function-based message authentication code / cipher-based message authentication code (HMAC/CMAC). key, KEK is then used for traffic encryption key (TEK) encryption and distribution. After this, the MS and the BS should perform TEK 3-way handshake. This means that the MS and the BS exchange the keys which are finally used for data traffic encryption, which are related to all security associations (SA) between them. Such an SA manages the keys for data encryption, their lifetimes and other security related parameters. A 3-way Handshake scheme is supported by Mobile WiMAX to optimize the re-authentication mechanisms for supporting fast handovers. The flow in SA-TEK 3-way handshake is shown in [7].

B. Mobile WiMAX Handover

Handover Management is the process of initiating and ensuring a seamless and lossless handover of a mobile terminal from the region covered by one base station to another base station. The BS associated with the mobile station before the handover is called the serving BS while the new BS is referred to as the target BS.

As mentioned before, IEEE 802.16e defines three types of handover: HHO, MDHO, and FBSS.

1) *HHO*: here, the MS communicates with just one BS in each time. In a HHO, the mobile station interrupts the communication with the serving BS and makes a transition to the target BS. This kind of handover is often referred to as a break-before-make. The serving BS periodically broadcasts an advertisement management message (MOB-NBR-ADV) which is decoded by the MS that will seek initial network entry or handover to obtain information about the characteristics of the neighboring BSs, such as the number of neighbor BSs and their BSIDs. The MS then selects the target BS for the HO, performs ranging, association procedures, authentication and registers with the target BS. Through ranging, the MS can acquire the timing, power and frequency adjustment information of the target BS. The main advantage of the HHO method is that it is the simplest of all the three, but it has high latency which is typically on the order 100 ms or more [11]. Thus, HHO is typically used for data services.

2) *MDHO*: in this scheme, the MS and BS maintain a list of BSs that are involved in the handover procedure. This set is called the Diversity Set (sometimes called "Active Set"). The MS continuously monitors the BSs in the diversity set and defines an "Anchor BS", which is one of the BSs from diversity set. Further, the MS is registered to the Anchor BS. In MDHO, the BSs are required to share MAC context, which includes current encryption and authentication keys associated with the connections.

3) *FBSS*: here, the MS maintains a valid connection simultaneously with more than one BS. The MS continuously monitors the diversity set as in MDHO, does ranging, and maintains a valid connection ID with each of them. However, the MS communicates with only one BS (anchor BS) for all types of uplink and downlink traffic. When a change of anchor BS is required, the connection is switched from one base station to another without having to explicitly perform handoff signaling. BSs involved in FBSS are also required to share or transfer MAC context.

Both FBSS and MDHO offer superior performance to HHO, but they require that the BSs in the active set be synchronized and share network entry related information.

C. Security of IEEE 802.16e Handover Schemes

Due to the resource constraints on most MSs, it may be too expensive for the MS to re-authenticate every time it registers to another BS because the authentication protocol is based on a public key infrastructure (PKI). Since a connection is switched from one base station to another, establishing a new security association to the new point of attachment is needed. To avoid the latency associated with the security association re-establishment, many network architectures have chosen to simply transfer the link security keys from one base station to the next. As a result, all secret keys used before the handover will be reused after the handover. This creates the domino effect [3], which means that if the security of one BS is compromised, it can lead to the security compromise of all previous BSs (backward secrecy) and following BSs (forward secrecy).

D. Vulnerabilities in IEEE 802.16e

1) Unauthenticated Messages

Mobile WiMAX includes some unauthenticated messages. Their forgery can interrupt the communication between the MS and the BS. One of these messages is the neighbor advertisement message. The serving BS sends this message to announce the characteristics of neighbor BS to MSs seeking for handover. An adversary can forge this message and prevent MSs to handover to BSs, which might have better characteristics as their serving BS.

The adversary can also distribute wrong data about neighbor BSs or announce non existing BSs. To avoid this attack [12], the non-authenticated messages sent on the primary or basic management connection can easily be authenticated using a HMAC or CMAC digest.

2) Shared Keys within Group Members

The sharing policy of the secret keys among the BSs in the diversity set has a severe security flaw which creates the domino effect stated earlier. Thus, to avoid this problem the least privilege principle [3] should be applied.

III. PRE-AUTHENTICATION SCHEME FOR SECURE HANDOVER

The EAP methods do not offer a fast re-authentication feature, causing too much delay in executing the re-authentication. To reduce the delay associated with the SA re-establishment, solutions are expected to include handover keying, low-latency re-authentication, and pre-authentication. In [13], the authors have chosen to transfer the link security keys from one BS to the other. However, those approaches cannot guarantee the backward or forward secrecy.

In this section, we suggest a pre-authentication scheme proposed in [7], which results in the establishment of an AK in the MS and the target BS before handover. Thus, upon handover, they only need to perform SA TEK 3-way handshake and update the TEK. We assume that the AS knows the neighbor BSs of each BS.

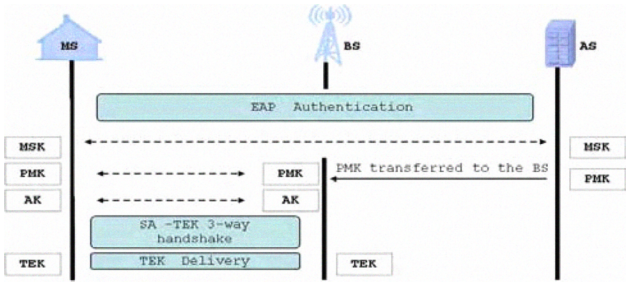

Fig. 2. Modified message exchange and key derivation

A. Hard Handover Pre-authentication Scheme

The MSK of an MS is delivered from the AS to the serving BS, as shown in Fig. 1. Any BS that received the MSK from the AS can derive the PMKs and AKs of other neighbor BSs using (1). Thus, for a secure pre-authentication, a unique key which binds the BSID and the MAC address of the MS should be generated by the AS and delivered to the corresponding BS instead of the MSK [7]. In the pre-authentication scheme as shown in Fig. 2, the PMK which binds the BSID and the MAC address of the MS is delivered to the corresponding BS. The PMK and AK are generated as follows:

$$\text{PMK} = \textit{Dot16KDF}(\text{MSK, MS MAC Address} \,/\, \text{BSID} \,/\, \text{“PMK”, 160}),$$
$$\text{AK} = \textit{PRF}(\text{PMK, 160}). \qquad (2)$$

where *PRF*(PMK, 160) is a cryptographically secure pseudorandom number function that generates an output of 160-bit length on the input of PMK.

In the pre-authentication phase, the AS generates unique PMKs and distributes them to the corresponding BSs, such that each BS receives a unique PMK which cannot be derived by anyone other than the MS and the AS. Then, the neighbor BSs derive their AKs for the MS. In the same way, the MS derives the PMKs and AKs for its neighbor BSs, as it knows the BSIDs of the neighbor BSs included in the MOB-NBR-ADV message. In Fig. 3, the pre-authentication procedure is shown. PMK*i* represents the PMK of a BS*i*. In the pre-authentication process, since the serving BS cannot know neither the MSK nor the PMK*i* of its neighbor BS*i*, it cannot derive the AK*i* of its neighbor BS*i*. Upon a handover, as the MS and the target BS already have the AK, only the 3-way handshake is performed and the TEK is updated [7].

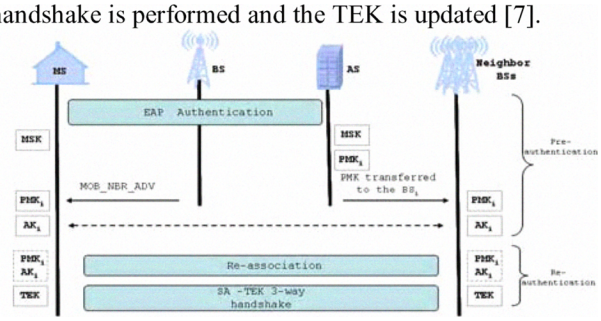

Fig. 3. Authentication procedure in the hard handover scheme introduced

B. Soft Handover Pre-authentication Scheme

In the soft handovers, BSs involved in MDHO or FBSS are required to share or transfer MAC context. The MAC context includes all information the MS and BS normally exchange during a network entry. Thus, the MS authenticated with one of the BSs in the diversity set is automatically authenticated with the other BSs from the same diversity set [7].

The sharing policy of the secret keys among the BSs in the diversity set has a severe security flaw which creates the domino effect stated earlier. Thus, to avoid the problem of domino effect, the least privilege principle should be applied where each entity gets the minimum amount of security information to continue its operation.

The FBSS pre-authentication phase is similar to the HHO pre-authentication, however, in the FBSS the AS is required to generate a unique PMK for each BS in the diversity set using (2) and distributes it to each of them. Then, with BSs in the diversity set, the MS only performs 3-way handshake and receives the TEKs from each of the BSs. While, in MDHO, a strong security assumption that all entities in the diversity set trust each other.

IV. FORMAL ANALYSIS OF IEEE 802.16E PKMV2 HANDOVER

In this section, we formally verify the IEEE 802.16e handover schemes and the pre-authentication protocol for both hard and soft handovers of using the Scyther tool [8]. Scyther is an automated security protocol verification tool, which combines the possibility of verification and falsification. It can verify protocols with unbounded number of sessions, with guaranteed termination. The tool also analyzes infinite sets of traces in terms of patterns, and supports multi-protocol analysis. It is the only currently existing tool capable of verifying synchronization. The operational semantics of Scyther is based on [14].

In order to systematically analyze the characteristic properties that define the essence of an attack, we first conduct an analysis of the main concepts of security protocols. Using the global description in [14], we can identify the following components of the security protocol model.

- *Protocol specification.* It describes the behavior of each of the roles in the protocol. Most often, a role in a security protocol is specified as a sequential list of events.
- *Communication model.* It describes how the messages are exchanged between the agents. We assume asynchronous communication with a single network buffer because this is the most general approach.
- *Agent model.* Agents execute the roles of the protocol. The agent model is based on a closed world assumption. This means that honest agents show only the behavior described in the protocol specification. The formal protocol specification describes the initial knowledge of agents required to execute a role and the declaration of functions, constants and variables occurring in the protocol specification, as well as the way the protocol should be executed.
- *Threat model.* Based on Dolev and Yao's network threat model [15] that the intruder has full control over the network, which means that the intruder is able to inject, block, alter, eavesdrop messages and anything that can be constructed can be inserted into the network. We also take into consideration the fact that there can be regular agents that have been compromised by the intruder. When an agent is compromised by the intruder, the intruder learns all the knowledge of this agent for all the roles.
- *Cryptographic primitives.* These are idealized mathematical constructs such as encryption, using the black box approach. This means that we will not consider the internal implementation of cryptographic primitives and we will only know their relevant properties.
- *Security requirements.* They are expressed as safety properties (i.e. something bad will never happen). In our semantics, we will only study secrecy and two forms of authentication (*non injective synchronization and non injective agreement)*

A. Protocol Model

We describe the behavior of the protocol in terms of its roles, either an initiator or a responder. Our system consists of four communicating agents, MS, BS, Authentication Server (AAA) and Neighbor Base Stations, specifically the target BS (BS_i). An agent can perform any number of roles in a protocol execution. A role performed by an agent is called a run. Agents execute their runs to achieve their security goals. While agents pursue their goals, an intruder may try to oppose them. In order to resist the intruder attacks, an agent can make use of cryptographic primitives when constructing messages.

The basic entities in our framework are role specifications that will be executed by agents. Every role specification consists of a sequence of events describing the messages the agent shall send and receive, as well as certain security claims. Every claim event in a trace results in a declaration about the trace that may or may not be true. We focus on two security properties in our work: secrecy and authentication. A secrecy claim event is essentially the statement that something is never known to the adversary's. Authentication is captured by the notion of synchronization this means, the claiming execution's events match read and send events of the protocol roles. Synchronization requires that corresponding send and receive messages have to be executed in the expected order.

The role for which the claim is tested is denoted by x and y is the message. The following claims are used

- *Claim (x,Secret,y)* the agent performing the role x knows that the intruder will never have knowledge of y
- *Claim (x,Niagree)* agent performing the role x knows that the message received is from an authenticated sender.

We will focus on the confidentiality of the keying material distributed. Confidentiality is the disclosure of the keying material to passive and active attackers of the key distribution protocol must not be possible. This claim is fulfilled if the authentication server has the guarantee that all exchanged keys (described as key) are secret. The formal definition of this property is given below

Property 1: $\forall_{key}$ *(claim (AAA, Secret, key))*

B. Formal Verification of Handover Scheme

To avoid the latency associated with the security association re-establishment during handover, many network architectures have chosen to simply transfer the link security keys from one base station to the next, which means all secret keys used before the handover will be reused after the handover. Thus, if the security of one base station is compromised, it can lead to the compromise of the security of all the following base stations. The formal verification of the handover scenario is shown, as follows:

→

- *AAA*→ *BS$_i$:* MSK
- *BS* *MS:* MOB_NBR_ADV

As explained earlier the AAA server will transfer the security keys from one base station to the next, thus the AAA server delivers the MSK to the target BS. Also, the serving BS will broadcast the MOB-NBR-ADV, which is an advertisement message decoded by the MS to obtain information about the characteristics of the neighboring BSs.

An adversary can eavesdrop the first message and obtain the MSK key sent to the target BS. This means that the security of the target BS is compromised, which can lead to the compromising of all previous and following BSs. As the identifications of the MS and BS are exposed to the public, the adversary can derive the PMKs and the AKs of the compromised BSs. By possessing the AK, attackers can decrypt the TEKs. Also an adversary can act as a serving BS (rogue BS), which opens the protocol to forgery attacks, where an unauthorized BS can communicate with a MS. In this case, the MS cannot decide whether the entity sending the AK is a legitimate BS or not. Any rogue BS can therefore forge a response message so that a rogue BS can be used to perform a man-in-the-middle attack.

To avoid this problem, a so called principle of least privilege is applied where each entity gets the minimum amount of security information to continue its operation.

C. *Formal Verification of Pre-authentication Handover*

The formal verification of the pre-authentication handover scheme is shown as follows:

- *AAA*→ *BS$_i$:* PMK$_i$
- *BS* *MS:* MOB_NBR_ADV

For a secure pre-authentication, a unique key (PMK$_i$) is delivered from the AAA server to the corresponding BS. However, an intruder can eavesdrop the first message and obtain the unique PMK key which binds the BSID and the MAC address of the MS. Thus if this unique key is obtained by an adversary, only the security of the target BS is compromised, hence preventing what is called the domino effect. Therefore, compromised BSs cannot derive the PMKs and AKs of other BSs due to the secrecy of the MSK or the PMK and the one-way property of the *Dot*16*KDF* key generation function. Thus, for the adversaries, 2^{160} brute-force searches are needed to determine the AK, which is considered computationally infeasible. Moreover, the least privilege principle prevents the previous serving BS from guessing an AK of the following target BS, and prevents the target BS from guessing an AK of the previous serving BS.

The pre-authentication protocol can be fixed as follows. To prevent any intruder from capturing the PMK delivered from the AS to the target BS, we propose to include a PKI certificate for the target BS (BS$_i$). The PMK key is encrypted with the public key of the target BS pk(BS$_i$) and sent to the BS$_i$. A nonce (N_1) can be added for identification. The formal analysis and modifications of the pre-authentication handover scheme is shown:

- *AAA*→ *BS$_i$: {PMK$_i$, N_1 }pk(BS$_i$)*
- *BS* *MS:* MOB_NBR_ADV

V. CONCLUSION

In this paper, we analyzed different security aspects of handover schemes defined for IEEE 802.16e as well as their vulnerabilities. Afterwards, we used the Scyther tool to formally verify the handover security of the IEEE 802.16 standard. This verification showed some flaws in the used standard. For example, some messages which carry sensitive information without any authentication were found. If they are forged this can be dangerous for the system operation. In addition, the IEEE 802.16e specification does not define a pre-authentication scheme. Although we believe that this pre-authentication handover scheme would be of high value for a more secured protocol. Therefore, we formally verified the proposed pre-authentication handover scheme and proved that it is opposing the domino effect. Finally, we proposed changes on the pre-authentication protocol and verified that using Scyther tool, such proposed changes will increase the security of Mobile WiMAX significantly.

REFERENCES

[1] IEEE Std. 802.16-2004: IEEE Standard for Local and Metropolitan Area Networks Part16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE, 2004.

[2] IEEE Std. 802.16e-2005: IEEE Standard for Local and Metropolitan Area Networks Part16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems, IEEE, 2006.

[3] R. Mao, M. Nakhjiri, X. Dong, "Mobility Sensitive Master Key Derivation and Fast Re-authentication for 802.16m," IEEE C802.16m-07/029, February 2007.

[4] B. Aboba, "Fast handoff issues," IEEE-03-155r0-I, IEEE 802.11 Working Group, 2003.

[5] International Telecommunication Union, "General Characteristics of International Telephone Connections and International Telephone Circuits," ITU-TG.114, 1988.

[6] L. Zan, J. Wang, and L. Bao, "Personal AP Protocol for Mobility Management in IEEE 802.11 Systems," Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services pp. 418-425, San Diego, USA, July 2005.

[7] J. Hur, H. Shim, P. Kim, H. Yoon and N. Song, "Security Considerations for Handover Schemes in Mobile WiMAX Networks," IEEE Wireless Communications and Networking Conference, pp. 2531–2536, March-April 2008.

[8] C. Cremers, "The Scyther tool: Automatic verification of security protocols," http://people.inf.ethz.ch/cremersc/scyther/index.html.

[9] D. Stanley, J. Walker, B. Aboba, "Extensible Authentication Protocol (EAP) Method Requirements for Wireless LANs," IETF Request for Comments (RFC) 4017, March 2005.

[10] K. Kim, C. Kim, and T. Kim, "A Seamless Handover Mechanism for IEEE 802.16e Broadband Wireless Access," Computational Science, LNCS 3515, Springer, pp. 527-534, 2005.

[11] D. Johnston and J. Walker, "Overview of IEEE 802.16 Security," IEEE Security and Privacy Magazine, vol. 2, pp. 40-48, May-June 2004.

[12] A. Deininger, S. Kiyomoto, J. Kurihara and T.Tanaka, "Security Vulnerabilities and Solutions in Mobile WiMAX," International Journal of Computer Science and Network Security, vol.7, no.11, pp. 7-15, November 2007.

[13] S. Xu, M. Matthews, C. Huang, "Security issues in privacy and key management protocols of IEEE 802.16," Procs. of ACM Southeast Conference, pp.113–118, Melbourne, Florida, USA, March 2006.

[14] C. Cremers and S. Mauw, "Operational semantics of security protocols," Scenarios: Models, Transformations and Tools Workshop 2003, Revised Selected Papers, LNCS 3466, Springer, 2005.

[15] D. Dolev and A. Yao, "On the Security of Public Key Protocols," IEEE Transactions on Information Theory, vol. 29, pp. 198–208, 1983.