

# A Theoretical Model for Location Privacy in Wireless Personal Area Networks

Dave Singelee<sup>1</sup>, Ford-Long Wong<sup>2</sup>, Bart Preneel<sup>1</sup>, and Frank Stajano<sup>2</sup>

<sup>1</sup> ESAT-SCD-COSIC, Katholieke Universiteit Leuven — IBBT,  
3001 Heverlee-Leuven, Belgium,  
`dave.singelee@esat.kuleuven.be`,

<sup>2</sup> Computer Laboratory, University of Cambridge,  
CB3 0FD Cambridge, United Kingdom

**Abstract.** Location privacy is one of the major security problems in a Wireless Personal Area Network (WPAN). The use of temporary pseudonyms has been suggested by several authors to solve the problem. In this paper, we construct a formal model of location privacy for WPAN. This theoretical framework contains a formal definition of location privacy and models the access of an adversary to the communication channels from a set of oracles. This theoretical model can be used to analyze and evaluate location privacy-enhancing pseudonym schemes proposed in the literature.

## 1 Introduction

### 1.1 Tracking mobile users

One of the most important security problems in Bluetooth, and in Wireless Personal Area Networks (WPANs) in general, is *location privacy* [6, 12]. When two or more Bluetooth devices are communicating, the transmitted packets always contain the Bluetooth hardware address of the sender and the destination (or an identifier which is directly related to this address). When an attacker eavesdrops on the transmitted data, he knows the unique hardware addresses of these devices. As these addresses can often be linked to the identity of the user operating the mobile devices, this corresponds to a violation of the privacy of the user. An attacker can obtain data on the time and place a user is located, and use this information to his benefit. This should definitely be avoided, the user has to decide when his location is revealed and when not.

Even when a Bluetooth device is in *non-discoverable mode* (in this mode, it does not respond to inquiries of other devices) or in *non-connectable mode* (in this mode, it does not respond to page scans of other devices), an eavesdropper observing transmitted data can obtain the unique hardware address of the mobile device. To make things even worse, the attacker does not have to be physically close to the communicating devices, he can use a device with a stronger (directional) antenna (e.g., it is very easy to construct an antenna which can

intercept Bluetooth communication from more than one mile away [3,4]) or just place a small tracking device near the two mobile devices.

Tracking users of mobile devices can have serious consequences. E.g., without location privacy, a terrorist could be capable of discovering in which hotel (and even in which room) an important politician stays. This would certainly entail serious security problems. Another example of an attack is to track users on a specific location and use this information for location dependent commercial advertisements (e.g., a shop can send advertisements to everybody that is nearby). This location based service can be desirable in some cases, but the user should be able to decide when his location is revealed and when not. Receiving such commercial messages on a mobile device could be quite annoying (e.g., comparable to SPAM sent via email).

As long as unique and fixed identifying information is used somewhere in (the header of) a message or in the construction of a certain sequence or pattern, it can be abused by an attacker to track the mobile device. There is not really a need for fixed identifiers in Wireless Personal Area Networks, as it only causes privacy concerns. The use of temporary pseudonyms has been suggested to solve the problem, as we will discuss later in this paper.

## 1.2 Overview of this paper

The remainder of this paper is organized as follows. Section 2 describes the location privacy problem more in details and defines the two goals of location privacy-enhancing techniques: establishing untraceability and unlinkability. A brief discussion on the use of temporary pseudonyms and how they could solve the location privacy problem, is presented in Section 3. It also contains an overview of some techniques that were proposed in the literature. A formal model of location privacy for Wireless Personal Area Networks is proposed in Section 4. Finally, section 5 provides a final conclusion on the paper.

## 2 Defining the Location Privacy Problem

The use of a fixed identifier (or information that is directly related to it) in the header of a message, and/or using the fixed hardware address as the input of a certain procedure, results to location privacy vulnerabilities. It is important to define the exact problem one needs to solve. In the location privacy problem, one tries to prevent other parties from learning one's current or past location [2]. Note that location privacy is different than traditional requirements such as anonymity or unobservability [5,9].

More in detail, one wants to solve the following scenario. There are two mobile devices, called *A* and *B*, that want to communicate privately (let us assume that *A* starts the communication). We implicitly assume that both devices are personal devices, belonging to a specific user (this does not have to be the same user). *A* sends a message to *B* using a wireless communication technology (e.g.,

Bluetooth). Such a message consists of a *header* and a *payload*. The header contains identification information (typically the address of the sender and receiver or information that is directly related to these addresses), the payload just plain data (encrypted or not). We want to investigate how  $A$  can send a message to  $B$ , in such a way that  $B$  still knows the message was intended for him, but that an attacker (and any third party) has no information about the identity of  $A$  and  $B$ .

The goal of location privacy-enhancing techniques is to establish untraceability and unlinkability. These concepts can be informally defined as follows:

- It should be computationally hard for an attacker, who observes the exchanged messages, to detect which specific device is participating in the communication. This property is called **untraceability**. Note that it is not a problem that an attacker detects a device is sending and/or receiving data, and that the attacker is even allowed to know the precise location of this device (e.g., by observing the signal strength of the radio transmission). However, the attacker should not be able to determine the exact identity (i.e. the unique hardware address) of this device.
- It should be computationally hard for an attacker to link several messages to one sender and/or receiver (even without knowing the exact identity of this device). This property is called **unlinkability**. If one can detect when a certain (unknown) device is communicating, one could maybe use this information to discover the unique hardware address of the device (e.g., by observing certain specific communication patterns) and hence track it. Note that unlinkability covers untraceability, but not vice versa.

In the design of location privacy-enhancing techniques, one typically assumes that the attacker is omnipresent, has significant computational resources (but is computationally bounded), and is able to mount active attacks. The adversary is hence able to perform active attacks such as replay attacks or inserting dummy traffic. The communication range of the attacker is not limited, as he can modify the antenna of his device to intercept communication from a large distance.

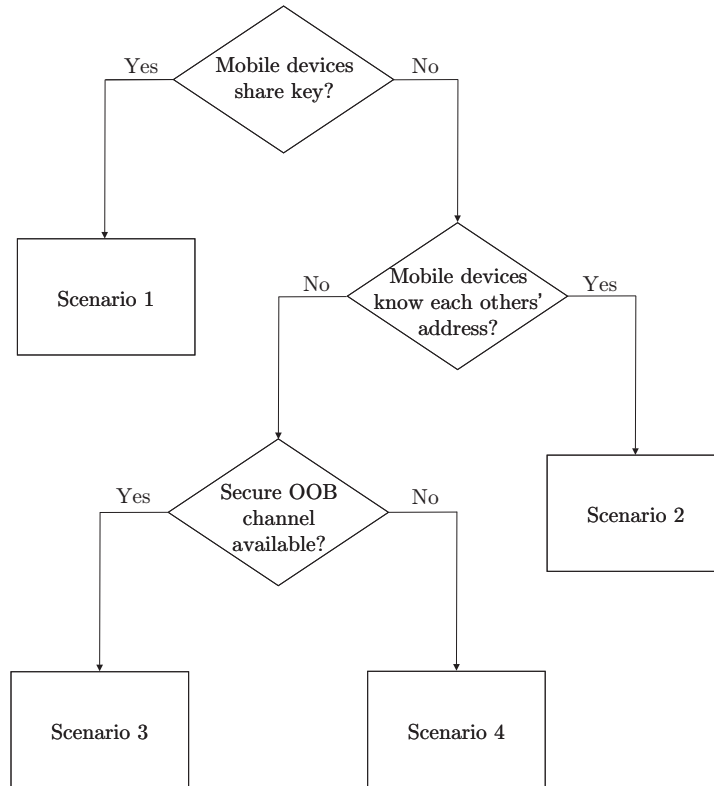
### 3 Using Temporary Pseudonyms

The location privacy problem in Wireless Personal Area Networks can be solved by using temporary pseudonyms instead of fixed identities. It is important that these pseudonyms are not completely stateless. Otherwise, pairing information, relationships between the different mobile devices and network configurations would be lost every time the pseudonym is updated. This would require a lot of re-initializations, which is definitely not efficient and user-friendly. Traditional pseudonym systems [8] cannot be employed, as one cannot make use of a central trusted server in a WPAN.

The mobile devices themselves have to make sure that location privacy is ensured. They will use shared data to compute a temporary pseudonym that replaces the fixed identifier in the header of the message. This random pseudonym,

which certainly has to be variable, will appear as random data for an eavesdropper, but the other party will recognize it and hence know the message was intended for him.

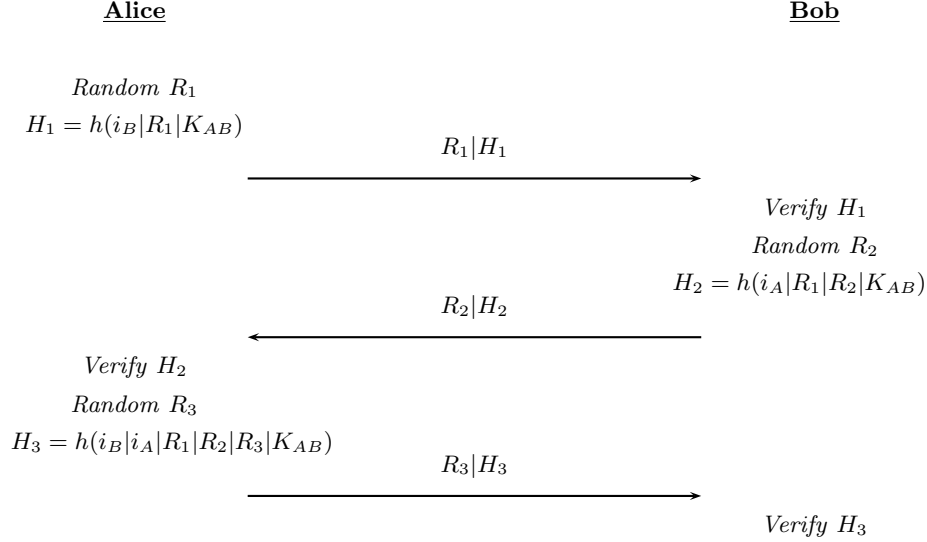
At least two solutions for location privacy in WPAN using temporary pseudonyms have been proposed in the literature. Singelée and Preneel [11, 13] gave an overview of four communication scenarios that could take place in a Wireless Personal Area Network, as depicted in Fig. 1. For each of these scenarios, they proposed an appropriate solution, which makes use of temporary pseudonyms, to create location privacy. They also demonstrated that the first two scenarios are basic scenarios, and that the other two scenarios can be converted (and hence reduced) to one of these basic scenarios. More details can be found in [11, 13].



**Fig. 1.** Four WPAN communication scenarios

Wong and Stajano proposed a protocol to provide location privacy in Bluetooth networks [15]. Their protocol is shown in Fig. 2 and consists of a three-way handshake. The three messages in the protocol are denoted by  $ID_1$ ,  $ID_2$  and  $ID_3$ . The relevant past pseudonyms of Alice and Bob are denoted by  $i_A$  and  $i_B$ .  $h()$  is

a cryptographic hash function (the collision probability and the first and second preimage resistance of the output of the hash function must be low).  $R_1$ ,  $R_2$  and  $R_3$  are random nonces, and  $K_{AB}$  is a link key shared by Alice and Bob. Both parties keep a database of tuples containing their own temporary pseudonym, the pseudonym of the other party, and the shared link key.



**Fig. 2.** Wong and Stajano's location privacy protocol

The use of temporary pseudonyms helps to avoid location tracking. The security of the protocol depends on the randomness of the nonces, the irreversibility of the hash function and the secrecy of the shared link key. After the successful execution of the three-way protocol, both parties know they are communicating with the correct party. After having verified the correctness of the  $ID_2$  message, Alice has the assurance that Bob knows his own previous pseudonym, the previous pseudonym of Alice, and their shared key. After having verified the correctness of the  $ID_3$  packet, Bob has the assurance that Alice knows these same three things. The past temporary pseudonyms are protected from all third parties.

## 4 Theoretical Location Privacy Model

To evaluate and analyze location privacy-enhancing pseudonym schemes for WPAN, one needs a universal theoretical framework. Such theoretical models already exist for RFID (e.g., the theoretical model proposed by Avoine [1], by

Juels and Weis [7], or by Vaudenay [14]), but not yet for WPANs. We adapted these models, and incorporated the specific properties of a Wireless Personal Area Network. E.g., in RFID communication, there is always a reader and a tag, while in a WPAN the nodes have equal functionality. The result of applying these models in a different setting is a theoretical location privacy framework for WPAN, which contains a formal definition of (the different types of) location privacy, and models the access to the communication channels from a set of oracles. Note that we will only consider protocol-level location privacy issues. In the real world, there could be many possible side channels which enable an attacker to trace a particular user.

We will now discuss our theoretical model more in detail.

#### 4.1 Overview of the different entities

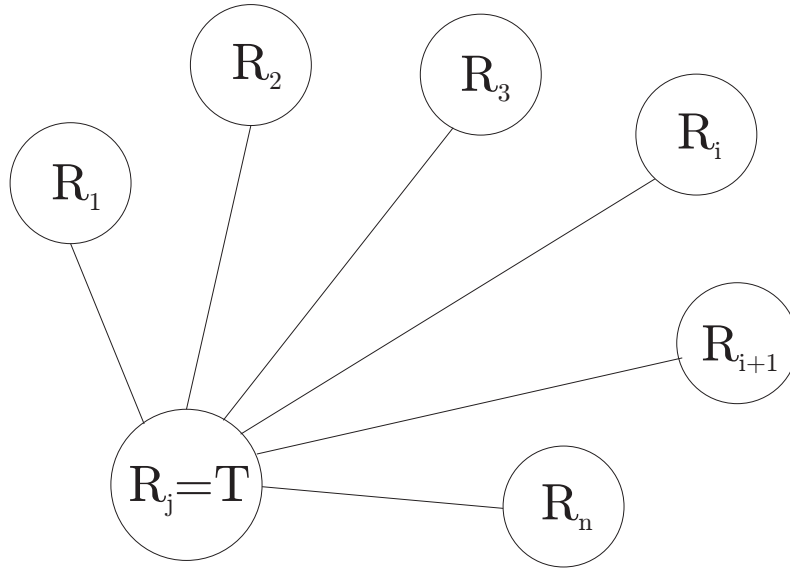
Before proposing a formal definition of location privacy in WPAN that can model a variety of security protocols and attacks, we need to define the different entities that appear in a system. A WPAN is formed by a group of mobile nodes  $R_i$ . Each of the nodes has equal functionality (in the sense that there is no client-server relation), they form a peer network. Typically, the WPAN contains a cluster of nodes that “intensively” communicate with each other, and always travel together in time and place. Such a cluster is called a *communicating constellation*. An example is the cluster of personal devices that a user carries with him every day (a mobile phone, PDA, watch, . . .). In the rest of this paper, we assume that all the devices in the communicating constellation are operated by the same user.

In the system, there is also an attacker present who wants to track a particular user by the devices the latter is carrying. In our theoretical location privacy model, this will be modeled by some attack games. An attack game always starts with the attacker being challenged. During this phase, the attacker chooses a particular node  $R_j$  (at random, or really a specific node). This node  $R_j$  is called the *target node*  $T$ . The goal of the adversary in an attack game is to distinguish between two different nodes, one of them being the target node  $T$ , within the limits of its computational power and taking into account other restrictions (related to the attack game). More information on the different attack games will be presented in Sect. 4.4.

The concept of a communicating constellation and a target node is depicted in Fig. 3. The attacker is not shown in this figure. Note that we assume that all nodes  $R_i$  know the node  $T$  (in the sense that they can recognize it during communication). Nodes that have never communicated with  $T$  before, are not interesting from an attacker point of view (they do not offer any new information), and are hence discarded.

#### 4.2 Identification protocol

During communication, nodes in a WPAN need to identify the source and destination of a message. There are several methods to do this. One can put the



**Fig. 3.** Communicating constellation in the WPAN

fixed identifier of a node in the address field in the header of a message. Of course, this causes privacy problems. One can also apply a more advance location privacy-enhancing protocol which uses pseudonyms instead of the fixed hardware addresses.

In our theoretical model, the protocol used to identify the source and destination of a message is modeled as an *identification protocol*  $P$ . Such a protocol  $P$  is always conducted between two nodes of the network. Each of the nodes  $R_i$  can however run several instances of  $P$ . In each round of  $P$ , one of the nodes initiates the communication, the other responds. To model the initiator and the responder, we can use the abstract messages “*start protocol*” and “*stop protocol*”. When a node receives the message “*start protocol*”, it will take the role of the initiator. When the last message of the protocol  $P$  is sent to a certain node, this node will reply with the abstract message “*stop protocol*”. By using such abstract messages, there will always be a message going to a node, and a response going back.

In our theoretical model, we make no assumptions on which entity takes the role of the initiator, neither about the number of messages in the protocol  $P$ . The consequence is that our theoretical model can be applied on a large set of protocols.

### 4.3 Adversarial model

A theoretical framework of location privacy requires a formalization of the adversarial model. Such a model consists of the means of the adversary and his goals. The means of the attacker are represented using the following oracles:

- **Query target  $T$** : The attacker sends a message to  $T$ , and observes the response.
- **Query node  $R_i$** : The attacker sends a message to  $R_i$ , and observes the response.
- **Execute  $(R_i, R_j)$** : The attacker forces  $R_i$  and  $R_j$  to perform a complete round of the identification protocol  $P$ , and eavesdrops on the messages sent between the two nodes. One of these nodes can be the target node, but this is not necessary.
- **Reveal node  $(T, t_{rev})$** : By employing this oracle, the attacker obtains the entire content of the memory of  $T$  at time  $t_{rev}$ . This oracle can only be used once and the other oracles can no longer be used on node  $T$  after time  $t_{rev}$ .

During an attack game, the attacker is allowed to make a particular number of queries to each (or some) of the oracles. We parameterize the number of *Query target* messages by  $qt$ , the number of *Query node* messages by  $qr$  and the number of *Execute* messages by  $qe$ . An adversary with these means is denoted by  $\mathcal{A}[qt, qr, qe]$ . The more queries an attacker is allowed to make, the more powerful he is. It is interesting to note that one *execute query* is equivalent to  $m$  consecutive *query node messages*, with  $m$  denoting the number of messages in the identification protocol  $P$ .

### 4.4 Attack games

We will now define several parameterizable attack games. The goal of an adversary in an attack game is to distinguish between two nodes of the WPAN, one being the target node  $T$ , within the limits of his computational power and not exceeding the number of allowed queries to the oracles presented above. To analyze the security of an identification protocol  $P$ , we assume that its security level can be parameterized by a security parameter  $k$ . We will use the notation  $poly(k)$  to represent any polynomial function of  $k$ .

**Attack game 1** The goal of this attack game is to distinguish between a specific target  $T$ , chosen by the attacker, and another random node. The attack game goes as follows:

1. The attacker selects a specific node  $R_j = T$  from a particular communicating constellation. This will be the target node for the challenge.
2. The attacker can query the three oracles (*Query target  $T$* , *Query node  $R_i$* , and *Execute  $(R_i, R_j)$* ), as described in Sect. 4.3. The number of allowed queries to these oracles are parameterized by  $qt$ ,  $qr$  and  $qe$  respectively.



3. The adversary selects two nodes,  $T_0$  and  $T_1$ . One of these nodes is equal to the target  $T$ , the other node is a random node  $R_x$ . The goal of the attacker is to indicate which one of these two nodes  $T_b$  is the target node  $T$ .
4. The attacker can query the three oracles (*Query target*  $T_i$ , *Query node*  $R_i$ , and *Execute* ( $R_i, R_j$ )), as described in Sect. 4.3. The number of allowed queries to these oracles are parameterized by  $qt$ ,  $qr$  and  $qe$  respectively.
5. The attacker has to decide which node  $T_b$  (so  $T_0$  or  $T_1$ ) is equal to the target  $T$ . The attacker wins when his guess of the bit  $b$  was correct.

**Definition 1 ((qt, qr, qe)-location privacy)** *A protocol  $P$  executed in a WPAN with security parameter  $k$  is (qt, qr, qe)-location private if:*

$$\forall \mathcal{A}[qt, qr, qe] : Pr(\mathcal{A}[qt, qr, qe] \text{ wins attack game 1 by guessing } b) \leq \frac{1}{2} + \frac{1}{|\text{poly}(k)|}$$

**Attack game 2** The goal of this attack game is to detect that a certain node belongs to a specific communicating constellation. The attacker does not want to make a distinction between the nodes in the communicating constellation, detecting that a node is part of the group is already enough. This attack makes sense from a practical point of view, since an attacker is typically not interested in detecting a specific device, but the user operating the device. And since a user is often carrying the same devices, which form a communicating constellation, this attack is sufficient to track the user.

The game goes as follows:

1. The attacker selects a particular communicating constellation, formed by the group of nodes  $R_i$ . This group is the target of the attacker.
2. The attacker can query the two oracles *Query node*  $R_i$  and *Execute* ( $R_i, R_j$ ), as described in Sect. 4.3. The number of allowed queries to these oracles are parameterized by  $qr$  and  $qe$  respectively.
3. The adversary (randomly) selects one of the nodes  $R_i$ . This node is removed from the communicating constellation. The attacker also selects another node, which is not part of the communicating constellation (and hence not known by the nodes  $R_i$ ). These two nodes are randomly defined as  $T_0$  and  $T_1$ . The goal of the attacker is to indicate which one of these two nodes  $T_b$  belongs to the communicating constellation (and is hence known by the other nodes  $R_i$ ).
4. The attacker can query the three oracles (*Query target*  $T_i$ , *Query node*  $R_i$ , and *Execute* ( $R_i, R_j$ )), as described in Sect. 4.3. The number of allowed queries to these oracles are parameterized by  $qt$ ,  $qr$  and  $qe$  respectively.
5. The attacker has to decide which node  $T_b$  (so  $T_0$  or  $T_1$ ) belongs to the communicating constellation formed by the nodes  $R_i$ . The attacker wins when his guess of the bit  $b$  was correct.

**Definition 2 ((qt, qr, qe)-constellation location privacy)** *A protocol  $P$  executed in a WPAN with security parameter  $k$  is (qt, qr, qe)-constellation location*

private if:

$$\forall \mathcal{A}[qt, qr, qe] : \Pr(\mathcal{A}[qt, qr, qe] \text{ wins attack game 2 by guessing } b) \leq \frac{1}{2} + \frac{1}{|\text{poly}(k)|}$$

**Relation between the attack games** Since distinguishing between two nodes of the WPAN is a stronger requirement than detecting that a certain node belongs to a particular communicating constellation, we have the following relation between the two attack games:

$$\text{Game1} \Rightarrow \text{Game2} \quad (1)$$

In other words, a protocol  $P$  that is  $(qt, qr, qe)$ -location private is also  $(qt, qr, qe)$ -constellation location private.

#### 4.5 Forward security

Since the mobile devices in a WPAN can easily get lost or stolen, or affected by a virus, it is important to incorporate *forward security* in our theoretical model of location privacy. A protocol is forward secure if an attacker who obtains the memory content of a mobile device (and hence the current secret keys and identifiers), is not able to track it in the past. The notion of forward security results in the following attack game.

**Attack game 3** The goal of this attack game is to distinguish between a specific target  $T$ , chosen by the attacker, and another random node, somewhere in the past. The attack game goes as follows:

1. The attacker selects a specific node  $R_j = T$  from a particular communicating constellation. This will be the target node for the challenge.
2. The attacker can query the four oracles (*Query target*  $T$ , *Query node*  $R_i$ , *Execute*  $(R_i, R_j)$ ), and *Reveal node*  $(T, t_{rev})$ , as described in Sect. 4.3. The number of allowed queries to first three oracles are parameterized by  $qt$ ,  $qr$  and  $qe$  respectively. The adversary is only allowed to make one *reveal* query on the target node  $T$ .
3. The adversary selects two nodes,  $T_0$  and  $T_1$ . One of these nodes is equal to the target  $T$ , the other node is a random node  $R_x$  of the communicating constellation. The goal of the attacker is to indicate which one of these two nodes  $T_b$  is the target  $T$ , at a particular time before  $t_{rev}$ .
4. The attacker can query the three oracles (*Query target*  $T_i$ , *Query node*  $R_i$ , and *Execute*  $(R_i, R_j)$ ), as described in Sect. 4.3. These queries are only allowed to take place at times  $t_i$ , where  $t_i < t_{rev}$ . The number of allowed queries to these three oracles are parameterized by  $qt$ ,  $qr$  and  $qe$  respectively.
5. The attacker has to decide which node  $T_b$  (so  $T_0$  or  $T_1$ ) is equal to the target  $T$ , at a particular time before  $t_{rev}$ . The attacker wins when his guess of the bit  $b$  was correct.

**Definition 3 ((qt, qr, qe)-forward location privacy)** A protocol  $P$  executed in a WPAN with security parameter  $k$  is  $(qt, qr, qe)$ -forward location private if:

$$\forall \mathcal{A}[qt, qr, qe] : \Pr(\mathcal{A}[qt, qr, qe] \text{ wins attack game 3 by guessing } b) \leq \frac{1}{2} + \frac{1}{|\text{poly}(k)|}$$

## 5 Conclusions

Location privacy is one of the major security problems in a Wireless Personal Area Network. The leakage of the device's unique hardware address enables an attacker to keep track of the place and time a mobile device is communicating. The hardware address of the device can often be linked to the identity of the user operating the mobile device, and this causes severe privacy problems. While the basic location privacy problem of using a long-term device address can be resolved by using temporary pseudonyms, an incomplete solution can give rise to linkability.

In this paper, we have constructed a formal model of location privacy for WPAN. This theoretical framework contains a formal definition of (the different types of) location privacy and models the access of an adversary to the communication channels from a set of oracles. This theoretical model has been used to analyze and evaluate several location privacy-enhancing schemes proposed in the literature. This resulted in several design flaws being discovered, as will be published in [10].

**Acknowledgments.** This work is partially funded by a research grant of the Katholieke Universiteit Leuven for D. Singelée, by the Concerted Research Action (GOA) Ambiorics 2005/11 of the Flemish Government, by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy) and by the Flemish institute IBBT.

## References

1. G. Avoine. Adversarial Model for Radio Frequency Identification. Cryptology ePrint Archive, Report 2005/049, 2005. <http://eprint.iacr.org/>.
2. A.R. Beresford and F. Stajano. Location Privacy in Pervasive Computing. *IEEE Pervasive Computing*, 3(1):46–55, 2003.
3. H. Cheung. The Bluesniper Rifle. <http://www.tomsnetworking.com/Sections-article106.php>, 2004.
4. DEFCON. Computer Underground Hackers Convention. <http://www.defcon.org>.
5. A. Hevia and D. Micciancio. An Indistinguishability-Based Characterization of Anonymous Channels. In *Proceedings of the 8th Privacy Enhancing Technologies Symposium (PETS '08)*, Lecture Notes in Computer Science, LNCS 5134, pages 24–43. Springer-Verlag, 2008.
6. M. Jakobsson and S. Wetzel. Security Weaknesses in Bluetooth. In *Proceedings of the Cryptographer's Track at the RSA Conference (CT-RSA '01)*, Lecture Notes in Computer Science, LNCS 2020, pages 176–191. Springer-Verlag, 2001.

7. A. Juels and S.A. Weis. Defining Strong Privacy for RFID. Cryptology ePrint Archive, Report 2006/137, 2006. <http://eprint.iacr.org/>.
8. A. Lysyanskaya, R. Rivest, A. Sahai, and S. Wolf. Pseudonym Systems. In *Proceedings of the 6th Annual International Workshop of Selected Areas in Cryptography (SAC '99)*, Lecture Notes in Computer Science, LNCS 1758, pages 184–199. Springer-Verlag, 1999.
9. A. Pfitzmann and M. Köhntopp. Anonymity, Unobservability, and Pseudonymity – A Proposal for Terminology. In *Proceedings of the International Workshop on Design Issues in Anonymity and Unobservability*, Lecture Notes in Computer Science, LNCS 2009, pages 1–9. Springer-Verlag, 2001.
10. D. Singelée. *Study and Design of a Security Architecture for Wireless Personal Area Networks*. PhD thesis, Katholieke Universiteit Leuven, 2008. 244 pages.
11. D. Singelée and B. Preneel. Location Privacy in Wireless Personal Area Networks. In *Proceedings of the 5th ACM Workshop on Wireless Security (WISE '06)*, pages 11–18. ACM Press, 2006.
12. D. Singelée and B. Preneel. Review of the Bluetooth Security Architecture. *Information Security Bulletin*, 11(2):45–53, 2006.
13. D. Singelée and B. Preneel. Enabling Location Privacy in Wireless Personal Area Networks. Cosic internal report, Katholieke Universiteit Leuven, 2007.
14. S. Vaudenay. On Privacy Models for RFID. In *Advances in Cryptology - ASIACRYPT '07*, Lecture Notes in Computer Science, LNCS 4833, pages 68–87. Springer-Verlag, 2007.
15. F. Wong and F. Stajano. Location Privacy in Bluetooth. In *Proceedings of 2nd European Workshop on Security and Privacy in Ad hoc and Sensor Networks (ESAS '05)*, Lecture Notes in Computer Science, LNCS 3813, pages 176–188. Springer-Verlag, 2005.