2004－ＣＳＥＣ－26 (16)
2004／7／20

An Assessment of Wireless Location Privacy Risks in High Precision Positioning System

Leping Huang [†‡] Kaoru Sezaki [‡]

† Arco Tower 17F, 1-8-1, Shimomeguro, Meguro-ku, Tokyo, 153-0064, Japan
‡ The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan
E-mail: † leping.huang@nokia.com, ‡ sezaki@iis.u-tokyo.ac.jp

Abstract: The advance of ISM-band radio-based tracking systems (for example, wireless LAN-based tracking system) extends the application of location-based services (LBS), but it also threatens to allow the movement of users to be tracked whenever they are transmitting frames. Several protection methods based on periodical address updates have already been proposed. In this paper, we first evaluate the privacy risks in wireless communication, and especially in wireless LAN network. Through our analysis, we identified that new correlation attacks can defeat current periodical address updated based protection method when the accuracy of positioning system is high. Correlation attack is defined as methods, which utilize the correlation between the old and new addresses of the same node in location tracking. To combat such attacks, we propose the concept of a silent period. A silent period is defined as a transition period between the use of new and old pseudonyms, when a node is not allowed to disclose either the old or the new address. Through analysis, we find that a silent period should contain a constant period and a variable period. The effect of the constant period is to mix the spatial relation between the node's disappearing points and emerging points. The variable period mixes the temporal relation between the node's disappearing times and emerging times. We further propose a general framework to measure the location privacy, which takes into account the probabilities that users contribute to each other's privacy level, and we show how to use this framework to calculate the location privacy for a wireless system. Based on this framework, we propose two measures, called the geographical anonymity set (GAS) and the tracking accuracy (TA). Finally, we conclude with a plan for future work.

Keyword: Location privacy, privacy, positioning, anonymous communication, wireless LAN

1. Introduction

Recent technological advances in wireless location-tracking present unprecedented opportunities for monitoring the movements of individuals. While such technology can support many useful location-based services (LBSs), which tailor their functionality to a user's current location, privacy concerns might seriously hamper user acceptance.

Recently, the advance in ISM (Industrial, Science and Medical)-band radio-based location-tracking technology greatly extends the application of LBS. Owing to advances in signal-processing technology, tracking systems can determine the position of a signal sender precisely by measuring the radio signal strength indication (RSSI) and/or the time of arrival (TOA) of signals at multiple receivers. Recent research on Bluetooth-based positioning systems has achieved an estimation accuracy of 2 meters [1], while wireless LAN-based tracking systems have achieved an estimation accuracy of 1 meter [2][3]. However, such technological advances also reduce the location privacy of users, because tracking systems exist that can measure the position of users whenever they transmit, and that can achieve high accuracy without cooperation or approval from users.

There are currently several projects researching methods to protect users' location privacy when transmitting on the ISM band. For example, the Bluetooth Special Interest Group (SIG) identifies the continuous release of a node's permanent MAC address as a critical threat to a user's location privacy, and proposes an *anonymity mode* [4] as a solution to this problem. Gruteser et al. have proposed a disposable interface identifier method to protect the user's location privacy in wireless LANs. The main idea of these two approaches is to protect location privacy by periodically updating the node's MAC address. Although this may be sufficient for current tracking technology with a resolution of several meters, we believe that current proposals may not prevent nodes from being tracked as locating technology improves and nodes can be more accurately located. This is because, as a device can be positioned accurately, it will be possible to find a strong correlation between a trail left by an old address and that left by a new address. The old and new addresses can therefore be linked. There are many useful techniques for this type of tracking; for example Kalman filters or particle filters [5][6]. Such tracking becomes easier as the distance between nodes increases, as the speed of a device decreases, and as the accuracy with which a device can be positioned increases.

This paper discusses a method using a silent period to combat such a correlation attack. The remainder of the paper is organized as follows. Section 2 provides background information and earlier techniques in the field of location privacy protection. In Section 3, we first define the framework of the communications and tracking system to be used throughout this paper, and then evaluate the location privacy risks using such a system. In Section 4 we propose our privacy protection algorithm, which is based on the concept of a silent period. Two metrics, the geographical anonymity set (GAS) and tracking accuracy (TA), are then defined and analyzed. Finally, we conclude the paper in Section 7 with a discussion of proposed future work.

2. Background and Earlier Techniques

For some Bluetooth devices using the Bluetooth 1.2 standard, there is a need to prevent location tracking using Bluetooth MAC (BD_ADDR), channel access code (CAC), device access code (DAC), or a hopping sequence. Since all these parameters are determined by the device address, they can be used to perform different types of location tracking. In addition, each Bluetooth device has a user-friendly name that is given out upon request. This name then becomes an easy target for tracking a Bluetooth unit. A Bluetooth unit in anonymous mode combats location tracking by regularly updating its active device address. The active device address is used in all communications. Furthermore, units in anonymous mode shall always reply with the string "anonymous" when the user-friendly name is requested. After the standard authentication procedure, a node can show its own permanent (private) address to its authenticated pair. In anonymous communication, this method can be classified as the frequent pseudonym update method.

The provision of anonymity and pseudonymity is not new. In the 1980s, Chaum et al. [7] worked extensively to develop techniques for secure, untraceable electronic transactions over fixed networks. They introduced the novel *mix network*, which is a set of servers that serially decrypt or encrypt lists of incoming messages. These are sent out in a random order, in such a way that an attacker cannot correlate output message with input messages without the aid of mixer nodes (when several messages are passed simultaneously). However, this goal of real-time communication cannot be achieved because this approach requires several public key encryption and decryption operations, and an intentional time delay, to defeat a correlation attack. Other famous systems in the area of anonymous communication include Onion Routing [19], Crowds [18], and Anonymizer [20]. Most of these try to provide either sender anonymity, or unlinkability between sender and receiver. To evaluate the privacy provided by those algorithms, several measures have been proposed. From these, the size of the anonymity set defined by Chaum [17] is one of the most widely used to measure the anonymity of the Dining Cryptographer's (DC) network. The anonymity set is defined as the set of participants who may have sent a particular message, as seen by a global observer who has also compromised a set of nodes. Recently, Serjantov et al. [16] and Diaz et al. [21] have proposed an information theoretic model independently to measure the degree of anonymity of such a system.

To protect location privacy, Beresford and Stajano have proposed the concept of the *mix zone* [10] based on Chaum's *mix network*. They assumed LBS application providers are hostile adversaries, and suggested that application users hide their own ID from providers. A mix zone for a group of users is defined as the largest connected spatial region in which none of the users in the area have registered an application callback. Because application providers do not receive any location information when users are in a mix zone, their identities are "mixed". Beresford and Stajano also point out a problem with the frequent pseudonym update method when the spatial and temporal resolution of the location-tracking system is high. They claim that it will be possible to find a strong correlation between a trail left by an old address and that left by a new address, as a device can now be positioned accurately. The old and new address can therefore be linked to defeat the frequent pseudonym update approach. They propose using the mix zone when a pseudonym is updated. Assuming users change their identity to a new and unused pseudonym whenever they enter a mix zone, applications that see a user emerging from the mix zone cannot distinguish that user from another who was in the mix zone at the same time. Therefore, the application provider cannot link people going into the mix zone with those coming out of it. The mix zone works well to protect a user's location from an application provider, but it is doubtful whether this technique is also effective when the attacker is a malicious eavesdropper of the communication channel. Unlike an application provider, an eavesdropper tracks user's movement by locating where a user transmits a frame, instead of reading the location information contained inside the frame. From an eavesdropper's point of view, location information is disclosed continuously whenever the user is transmitting frames. If the

eavesdropper is instead considered the adversary in the mix zone, the mix zone now only includes areas where no node sends any frames. In wireless communication, this requirement means it is necessary to deploy anechoic chambers to restrict users' communication with external nodes. It is infeasible to deploy as many chambers as required to provide location privacy protection.

Recently, Gruteser and Grunwald have worked extensively on protecting location privacy in wireless LANs. They present a middleware architecture and algorithm to adjust the resolution of location information along spatial and temporal dimensions [12], and enhanced location privacy by frequently disposing of a client's interface identifier [1]. They propose updating the node's interface identifier whenever a station associates to a new access point. From their experiment, a node associates with a new access point once every approximately 30 minutes. They assume that an attacker may comprise some of the access points, and may track a user's movements based on the information with which the access point user is associated. Location-tracking methods that measure the RSSI and/or the TOA of frames are not considered attacking methods in their paper. However, RSSI/TOA-based tracking methods have much higher tracking accuracy than those that only analyze association logs in the access point (AP), and are stronger threats to location privacy. We should not ignore such an attack when providing location privacy protection.

3. Framework and Privacy Analysis

3.1. System Model

The system described in this paper is composed of four types of node: authentication server (AS), access point (AP), station (STA), and eavesdropper (E). Access point, station and eavesdropper nodes are incorporated with wireless LAN radio interfaces operating at identical frequencies. In commercial hotspot wireless LAN services, users of STA nodes always contract with one service provider. This service provider controls at least one AS. The area around users may be covered by other APs not controlled by the station's contracted service provider. The eavesdropper is capable of working in sniff mode, where it can capture all frames transmitted in the channel within its proximity. In addition, it is assumed that the wireless LAN interface in the eavesdropper is capable of providing a radio signal strength indication (RSSI) or time of arrival (TOA) for the eavesdropper's upper layer to estimate the STA's current position. We also assume that all regions that the STA may visit are covered by at least three eavesdroppers.

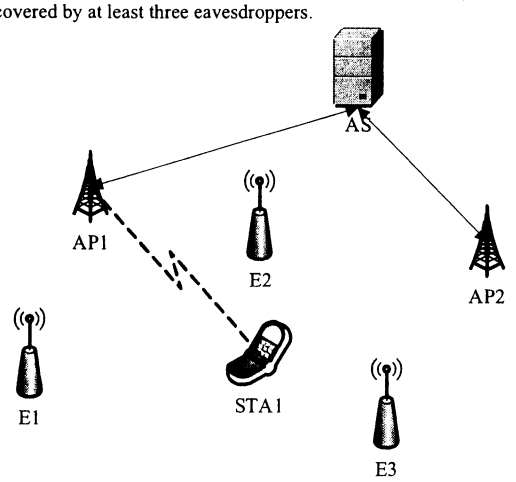

Figure 1. System Architecture

We illustrate the system architecture in Figure 1. Before data communication, STA1 first authenticates with AP1 by utilizing the authentication information stored in AS. After authentication, the STA communicates with the external network via AP1. When STA1 moves to the area covered by AP2, it should reauthenticate with AP2 using AS. STA1 is then within the proximity of three eavesdroppers (E1, E2, and E3). Eavesdroppers monitor the movement of the STA continuously by measuring the RSSI and/or the TOA information of each frame sent by STA1.

3.2. Privacy Risks for Wireless LANs

Generally, a location privacy threat occurs when an untrusted party can locate a transmitting device and identify the subject using the device. A wireless LAN network poses very serious location privacy threats for the following reasons.

Shared channel radio The MAC layer of 802.11 is based on CSMCA/CD, which requires all stations in its proximity to share the same communication channel. Any node can overhear all frames sent by others within its proximity.

Unsecured frame header The format of the 802.11 MAC frame is shown below in Figure 2. The encryption algorithms of 802.11 (AES and WEP) provide data confidentiality only for the frame body field of the frame; other fields of the frame are sent in plain text. Considering the shared channel characteristics of 802.11 radios, it is obvious that the station understands the sender address (field A1) and receiver address (field A2) of all frames sent by nodes within its proximity. This opens the possibility for an adversary to monitor the movement of a station without its cooperation.

Frequent broadcast of MAC address An eavesdropper can capture the MAC address of all frames sent in its proximity. If the sender transmits frames continuously, an eavesdropper can obtain the identity of the sender regularly. Continuous reception of the frame identity improves the accuracy of the tracking system. Here, we calculate the highest MAC address broadcast frequency by calculating the time to exchange one IP frame (containing only an IP header) between the AP and one station. The operation rate of 802.11 used here is 11Mbps DSSS/CCK in short-preamble mode. The procedure of exchanging one frame is as follows. The station first sends an IP frame (containing only an IP header), and the AP sends back an acknowledgement after waiting for the SIFS (short inter-frame space) after reception to avoid frame collision.The station then waits for the SIFS again before being ready for the next transmission. The duration of sending one frame is as follows.

$$\text{Time} = T_{PPDU\_IP} + SIFS + T_{PPDU\_ACK} + SIFS$$
$$= (T_{PPDU\_Header} + T_{PSDU\_IP}) + SIFS + (T_{PPDU\_Header} + T_{PSDU\_ACK}) + SIFS$$
$$= (96us + 54*8/11us) + 10us + (96us + 14*8/11us) + 10us$$
$$\cong 262us$$

Therefore, the eavesdropper will receive the identity of sender at a frequency up to 5KHz. In general, more frequent identity broadcasting improves the location-tracking accuracy.

2	2	6	6	6	2	4
Frame Control	Duration / ID	A1	A2	A3	Sequence Control	FCS
6	0–2312					
A4	Frame Body					

Figure 2. MAC Frame format

High tracking accuracy The radio signal properties of a WLAN system allow relatively precise determination of a client's position. When an access point receives a signal from a client, it is highly probable that the client's position will be within a typical range of the AP (say, 100 meters). By using triangulation methods based on the RSSI and/or TOA received by multiple cooperating access points, some systems can achieve very high accuracy *without cooperation* from the STA. Recent research reports an accuracy of up to 1 meter in an indoor environment [2].

Low-cost wireless LAN radio It is relatively inexpensive to

deploy enough wireless LAN nodes to cover large areas for location tracking, compared with covering the same area with a cell-based tracking system. This no technical reason also greatly increases the risk to location privacy in wireless LANs.

3.3. Attack Model

We assume that the clocks of eavesdroppers are synchronized. Eavesdroppers measure the RSSI/TOA of frames received from a specified STA to estimate its position. Frequent broadcasts of the MAC address provide adequate temporal and spatial tracking accuracy. STA uses the same periodical pseudonym update approach as that specified in the Bluetooth anonymity mode. With enough temporal and spatial precision, it may be possible for an adversary to correlate two pseudonyms that are sent separately from the same device moving through space. Temporal correlation may be used because the period with which stations change their pseudonym may be small. Spatial correlation may be used if it is assumed that a station will generally continue in the same direction, with the same speed as it traveled in the past. Such correlation attacks become easier as the distance between devices increases, as the speed of a device decreases, and as the accuracy with which a device may be located increases.

An example is shown in Figure 3. The definitions of disappearing time (DT), disappearing place (DP), emerging time (ET), and emerging place (EP) can be found in **Table 1**. Assuming that the system provides enough temporal and spatial precision, a node coming from the right with identity A updates its identity to A′ at some time within $[DT_1, ET_1]$. The last trail monitored by the system for A is recorded at time DT_1 and position DP_1; the first trail monitored by the system for A′ is recorded at time ET_1 and EP_1. From the similarity between both the times and locations of the two records, the tracking system can infer that node A changed its address to A′ sometime between $[DT_1, ET_1]$ and near the position $[DP_1, EP_1]$.

The objective of an adversary is to link A with A′ with high probability by using knowledge such as the tracking history for A, a user movement model, or a building layout.

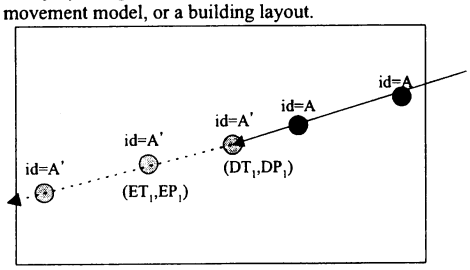

Figure 3. The correlation attack on periodical pseudonym updates

4. Proposal: Silent Period

To combat the correlation attack discussed in the previous section, we propose the new concept of a *silent period*. The silent period is defined as a transition period between using new and old pseudonyms in which a station is not allowed to disclose either the old or the new pseudonym. As a result, the silent period introduces ambiguity into the determinations of the time and/or place at which a change of anonymous address occurred. This makes it more difficult to associate two separately received pseudonyms with the same station, because the silent period disrupts the temporal and/or spatial correlation between two separately received pseudonyms and obscures the time and place where a pseudonym changed.

When multiple stations within the same region follow the rule of the silent period by ceasing transmission after updating their MAC address (pseudonym), the effect is the same as if those nodes had entered a mix zone in which no user's movement could be monitored by the system. Therefore, the silent period can be seen as an extension or implementation of the mix zone concept. It creates virtual mix zones by controlling the transmission of frames.

An example using the silent period is shown in Figure 4. In the figure, node 1 moves along a path from the upper right corner to the lower left corner. Meanwhile, node 2 moves along a path from the lower right corner to the upper left corner. Both nodes update their addresses and then enter the silent period. The effects of the proposal are illustrated near the intersection of the path of both nodes. The silent period is illustrated as a rectangle in the center of this graph. Node 1 arrives at the border of the silent period at time DT1 and position DP_1, and node 2 arrives at time DT2 and position DP_2. Here we assume that two nodes arrive at the border simultaneously ($DT_1 = DT_2$). Both node 1 and node 2 disable frame transmission for a silent period. After the silent period, nodes 1 and 2 restart frame transmission. The tracking system monitors a new trail with address A′ emerging at position EP_1 and time ET_1, and another new trail with address B′ emerging at time ET_2 and position EP_2. The tracking system knows that node 1 changed its address within the silent period. However, as it detects that two new address A' and B' emerge after the silent period, the tracking system cannot determine whether node 1 changed its address to A′ or B′, because both position EP_1 and EP_2 are reachable by node 1 from position DP_1, due to the silent period introduced during the moving time. This method obscures the temporal and spatial correlation between new and old pseudonyms by "mixing" the pseudonyms of nodes.

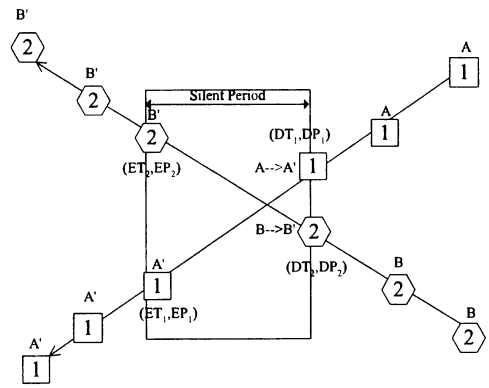

Figure 4. Illustration of the movement of two stations, node 1 and node 2, which both use a silent period

The discussion above assumes that nodes 1 and 2 switch their addresses at the same time for the sake of simplicity. The length of the silent period is assumed to be implicitly constant and identical for all nodes. However, when the address update time is not synchronized between nodes, the constant length of the silent period may not effectively mix the pseudonyms of nodes. This is because when the length of the silent period is constant, the tracking system can link the old and new pseudonyms to the same node. It does this by comparing the order of the emerging times of the two nodes (with their new pseudonyms) with the disappearing time of those with the old pseudonyms. As shown in Figure 4, if the tracking system knows that node 1 with identity A enters the silent period at a time DT_1 (earlier than node 2), it can infer that the emerging time of node 1 is earlier than that of node 2. When the system detects that a node with identity A′ emerges earlier than another node with identity B′, the system can easily infer that node 1 updates its address to A′.

To solve this problem, we propose the use of a variable length silent period. If the range of the emerging time of a node overlaps with that of another node, the tracking system cannot link the

detected pseudonym to the correct node. We illustrate the idea of a variable-length silent period in Figure 5. The silent period is determined by a random variable varying within the range [$T2_{min}$, $T2_{max}$]. When the ranges of the emerging times of two nodes overlap, the temporal relation between the emerging time and the disappearing time of the two nodes is obscured. The tracking system cannot correctly link the new pseudonym of a node to its old pseudonym. This approach is motivated by the CSMA/CD medium-access control algorithm used in collision-avoidance schemes.

In summary, the silent period should contain constant and variable periods. The effect of the constant period is to mix the *spatial* relation between a node's disappearing points and its emerging points, while the effect of the variable period is to mix the *temporal* relation between the node's disappearing times and its emerging times.

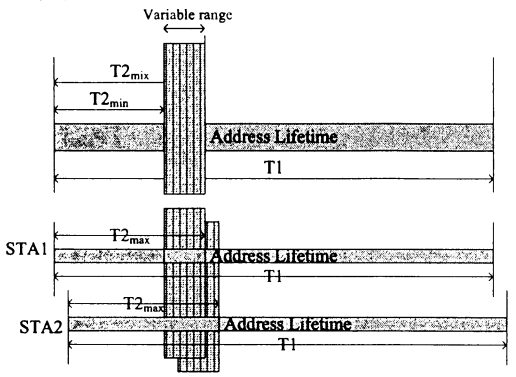

Figure 5. Synchronization issues in silent networks. Anonymity is guaranteed when the variable periods of T2 overlap

5. Location Privacy Measurement Framework

There are several factors that may influence the performance of the silent period approach. They are (1) the duration of the silent period, (2) the accuracy of the positioning system, (3) the mobility model of the individuals, (4) the density of users, and (5) the timing of address updates (i.e., synchronized or unsynchronized updates). In principle, longer silent periods and/or a higher density of individuals will improve privacy levels, as will more random movement of individuals. More accurate positioning systems cause lower privacy levels, as do unsynchronized pseudonym switches.

In this section, we first define a general framework to measure location privacy, using synchronized and unsynchronized address updates. In next section, we first evaluate the performance of our proposal by analyzing privacy level at some condition, and then show some preliminary simulation results.

We classify all nodes involved in the location-information protection system as two types, *target* and *mixer*. Target is the node whose privacy level is being measured. All other nodes that involved in this system are called as mixer. Mixers contribute to the privacy of target by restricting frame transmission for silent period of time to obscure the temporal and spatial relation between their and target's pseudonyms. The role of mixer and target may change depending on the object that tracking system is monitoring.

Our measurement framework is motivated by the information theoretic metric of anonymity independently proposed by Serjantov et al. in [16] and Diaz et al. in [21]. In their papers, they identified that not all nodes involved in anonymous communication contribute same degree of anonymity to the system. The size of anonymity set cannot precisely describe the degree of anonymity a system provides. In their proposals, they take into account the probabilities of user sending and receiving the message, and propose the use of an information theoretic measure, entropy of all users' probabilities of sending and/or receiving message, as the measure of anonymous system.

In location privacy protection system, we think that not all mixers contributes same level of privacy to *target*. Number of mixers that participates in the system can not precisely quantify the degree of privacy. Only those mixers, which arrive at the same area with target node and whose emerging time overlaps with that of target, contribute to the privacy level of target node. The contribution depends on the distribution of mixers and target's destination after silent period, and temporal relation between address update time of target and that of mixers.

The objective of our measures is to find out the *expectation* of privacy level contributed by all mixers over all reachable area of target, by considering the difference of contribution caused by spatial distribution of nodes' destination and temporal relation of node's emerging time. Let $PC_{x,y,t}$ be the discrete random variable of privacy contribution at position (x,y) and time t as shown in Figure 6. The values of $PC_{x,y,t}$ vary within a range of $\{\alpha_1,\alpha_2, \ldots, \alpha_i,\ldots\}$. Each value of privacy contribution α_i is only decided by the number of mixers, which arrive at small area [(x,y),(x+dx,y+dy)] after t seconds, and whose emerging time overlaps with that of target. We define two-variable function $p_{x,y,t}(i,j)$ as the probability mass function of $PC_{x,y,t}$. For a given position (x,y) and time t, Each value of $p_{x,y,t}(i,j)$ represents that i nodes that arrive at position after t seconds, and the emerging time of j nodes from those i nodes overlaps with that of target. In addition, we also define a one-variable probability mass function $r_{x,y,t}(i)$, while i represents each possible number of nodes that arrive at position (x,y) after t seconds no matter whether its emerging time overlap with that of target. The relationship between two variable probability function $p_{x,y,t}(i,j)$ and one variable probability function $r_{x,y,t}(i)$ is shown below for any given number of arrived node i.

$$r_{x,y,t}(i) = \sum_{j=0}^{i} p_{x,y,t}(i,j) \qquad (1)$$

$$p_{x,y,t}(i,j) = 0 \quad while \ j > i$$

In addition, we also define a continuous random variable Q_t, which represents target's destination distribution after t seconds. Two variable function q(x,y,t) is defined as the probability distribution function of Q_t over two dimensional geographical area.

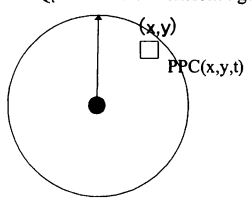

Figure 6: Illustration of privacy contribution formula

Before calculating the expected privacy level of target node, we first compute the expected privacy level at any given position. Here we define a new terminology position privacy contribution $PPC_n(x,y,t)$ as the contribution of privacy by all mixers of target n at position (x,y) after t seconds.

When the address update time of mixers are not synchronized with that of target, $PPC_n(x,y,t)$ is defined as below.

Definition 1: Position privacy contribution $PPC_n(x,y,t)$ (unsynchronized case)

$$PPC_n(x,y,t) = E\left|PC_{x,y,t}\right| = \sum_{i=0}^{N}\sum_{j=0}^{i}\alpha_j p_{x,y,t}(i,j) \quad (2)$$

$PPC_n(x,y,t)$ is the expected value of discrete random variable $PC_{x,y,t}$ at position (x,y) and time t for all possible combination of arrived node number i and overlapped node number j.

Definition 2: Position privacy contribution $PPC_n(x,y,t)$ (synchronized case),

$$PPC_n(x,y,t) = \sum_{i=0}^{N}\alpha_i r_{x,y,t}(i) \quad (3)$$

When the address update time of all mixer are synchronized with that of target, the emerging time of i nodes overlap with that of target. Consequently, the value of probability distribution function p(i,j) can be rewritten as below according to equation (1).

$$p_{x,y,t}(i,j)\begin{cases}0, & while\ i \neq j\\ r(i) & while\ i = j\end{cases}$$

Let us substitute p(i,j) in equation (2) with r(i) by using the relationship between r(i) and p(i,j) given above, equation (2) can be rewritten as equation (3) easily.

Definition 3: Node privacy level *NPL(t,n)*: Privacy level that node n receives contributed by node n's mixers involved in same location protection system.

From its definition, NPL is the expectation of Q_t over all reachable area of target. The value of Q_t at each position (x,y) are given by $PPC_n(x,y,t)$. Consequently, NPL can be given as below.

$$NPL(t,n) = \iint q(x,y,t,n)PPC_n(x,y,t)dxdy \quad (4)$$

Definition 4: System Privacy Level *SPL(t)*: the average of position privacy level that each node receives.

$$SPL(t) = \frac{1}{N}\sum_{n=1}^{N}NPL(t,n) = \frac{1}{N}\sum_{n=1}^{N}\iint q(x,y,t,n)PPC_n(x,y,t)dxdy \quad (5)$$

In the equations above, we only define that the value of privacy contribution α_i is related to number of arrived nodes. Here, we further propose the equation of privacy contribution based on number of nodes. Two measures are proposed. here The first one geographical anonymity set (GAS) is proposed to measure the degree of anonymity set in location privacy protection system. GAS can be seen as an extension of traditional metric of anonymous system, size of anonymity set. The second metric tracking accuracy (TA) is proposed to measure the accuracy of location tracking system when nodes use silent period algorithm. Details of these two measures are given below.

Geographical anonymity set The anonymity set of a node is defined as the set of all possible subjects that may be involved in anonymous communication. Many previous researches use the size of anonymity set as metric of systsm. For example, if there are ten nodes in a mix network, the size of the anonymity set is ten. As discussed above, such metric is not suitable for measuring location privacy in a large area, because each node may contribute different level of privacy to its different neighbors. However, we think that the size of anonymity set is still effective when the area is small enough and emerging time of all arrived node overlaps with that of target. If there are *i* mixer nodes (excluding the target) within area $[(x, x+dx), (y,y+dy)]$, the anonymity set within this area is $(i+1)$ according to traditional anonymous research. Consequently, the value of privacy contribution α_i is assigned the number of nodes including target arrived at this area simultaneously as below..

$\alpha_i = (i+1)$.

And the geographical anonymity set of a node when address update time is unsynchronized is given in Eq.(6) below, and that when address update time is synchronized is given in Eq. (7).

$$GAS(t,n) = \iint q(x,y,t,n)\left(\sum_{i=0}^{N}\sum_{j=0}^{i}(j+1)p_{x,y,t}(i,j)\right)dxdy \quad (6)$$

$$GAS(t,n) = \iint q(x,y,t,n)\left(\sum_{i=0}^{N}(i+1)r_{x,y,t}(i)\right)dxdy \quad (7)$$

Equation (6) and (7) is determined by the spatial probability distribution of mixers and target's destination, and temporal relation of address update timing between target and mixers. This gives us more precise description about the location privacy a system provides, than the number of nodes involved in location protection system.

Tracking accuracy (TA) The tracking accuracy of a node is defined as the probability that the tracking system can correctly link the old and new pseudonym of the node after the silent period. If there are *i* mixer nodes (excluding the target) within a small area $[(x, x+dx), (y,y+dy)]$, we think that the probability nodes can be correctly modeled with $1/(i+1)$. As mentioned in GAS, this assumption is only effective when the size of area is small enough. Consequently, tracking accuracy of a small area 1/(i+1) is assigned to privacy contribution α_i. We substitute α_i by tracking accuracy 1/(i+1) in Eq.(2), then the tracking accuracy of a node is defined as below.

$$TA(t) = \iint q(x,y,t)\left(\sum_{i=0}^{N}\sum_{j=0}^{i}\left(\frac{1}{j+1}\right)p_{x,y,t}(i,j)\right)dxdy \quad (8)$$

6. Conclusions and Future work

In this paper, we first evaluated the new risks imposed by high-accuracy location-tracking systems. A correlation attack is identified as a threat that cannot be defeated using existing periodical pseudonym update solutions. We proposed the new concept of a *silent period* to combat correlation attacks. Through analysis, we determined that the silent period should contain constant and variable periods. The effect of the constant period is to mix the *spatial* relation between a node's disappearing points and emerging points, while the variable period is used to mix the temporal relation between a node's disappearing times and emerging times. We also proposed a general framework to measure the performance of location privacy protection algorithms. This framework is suitable for both synchronized and unsynchronized address updates. We proposed two measures using this framework, the geographical anonymity set (GAS) and tracking accuracy (TA), and presented a preliminary evaluation of our proposal.

There are many open issues to study in the future. We should first conduct simulations to study the impact of length of silent period, node density, system accuracy on the privacy level. It is important to evaluate the impact of variable part of silent period on system privacy level when address update is not synchronized. We need to find out the relationship of threshold value with node density and system accuracy. It is also very important to study the integration problem with current wireless LAN protocol. Some issues, such as the impact on TCP/IP layer address duplication when proposed silent period method is used, should be studied further.

7. Acknowledgement

At the end of this paper, we would to thank for Prof. Kanta Matsuura for the excellent comments and useful discussion regarding to privacy and anonymous communication.

8. References

[1] Udana Bandara, Mikio Hasegawa, Masugi Inoue, and Hiroyuki Morikawa "Design & Implementation of a

Bluetooth Based Indoor, Location-Sensing System", Technical Report at the Third Meeting of IPSJ Ubiquitous Computing System WG, Tokyo, Japan, Feb. 2004.

[2] Hitachi: http://www.hitachi.co.jp/airlocation/

[3] P. P. Bahl, V.N. Padmanabhan, "RADAR: an In-Building RF-Based User Location and Tracking System", presented at INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, IEEE, 2000.

[4] Bluetooth SIG, "Bluetooth Standard 1.2 Draft 4", April 2003.

[5] Dieter Fox, Jeffrey Hightower, Lin Liao, Dirk Schulz, and Gaetano Borriello, "Bayesian Filtering for Location Estimation", IEEE Pervasive Computing, vol. 2, no. 3, pp. 24-33, IEEE Computer Society Press, July-September 2003.

[6] Dirk Schulz, Dieter Fox, and Jeffrey Hightower, "People Tracking with Anonymous and ID-Sensors using Rao-Blackwellised Particle Filters", the Eighteenth International Joint Conference on Artificial Intelligence (IJCAI), 2003.

[7] David Chaum, "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms", Communications of the ACM 4(2), February 1981.

[8] M. Gruteser, "Enhancing Location Privacy in Wireless LANs through Disposable Interface Identifiers: a Quantitative Analysis", 1st ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots, 2003.

[9] M. Gruteser and D. Grunwald, "A Methodological Assessment of Location Privacy Risks in Wireless Hotspot Networks", First International Conference on Security in Pervasive Computing, 2003.

[10] A. R. Beresford and F. Stajano, "Location Privacy in Pervasive Computing", IEEE Pervasive Computing, vol. 2, pp. 46–55, 2003.

[11] B. Schilit, J. Hong, and M. Gruteser, "Wireless Location Privacy Protection", pp. 135–137, IEEE Computer Magazine, Dec. 2003.

[12] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking", International Conference on Mobile Systems, Applications, and Services (MobiSys), CA, USA, 2003.

[13] A. Pfitzmann and M. Köhntopp, "Anonymity, Unobservability, and Pseudonymity: a Proposal for Terminology", International Workshop on Designing Privacy Enhancing Technologies: Berkeley, California, USA, 2001.

[14] D. Samfat, R. Molva, and N. Asokan, "Untraceability in Mobile Networks", in Proceedings of the 1st Annual International Conference on Mobile Computing and Networking: ACM Press, pp. 26–36, 1995.

[15] Dragoş Niculescu and Badri Nath, "Error Characteristics of Ad Hoc Positioning Systems", ACM MOBIHOC 2004, Tokyo, May 2004.

[16] A. Serjantov and G. Danezis, "Towards an Information Theoretic Metric for Anonymity", presented at Privacy Enhancing Technologies, 2002.

[17] D. Chaum, "The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability", J. Cryptol., vol. 1, pp. 65–75, 1988.

[18] M. Reiter and A. Rubin, "Crowds: Anonymity for Web Transactions", ACM Transactions on Information and System Security, vol. 1, 1998.

[19] Paul Syverson, Gene Tsudik, Michael Reed, and Carl Landwehr, "Towards an Analysis of Onion Routing Security", In the Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability, July 2000, pp. 96–114.

[20] Anonymizer: http://www.anonymizer.com/.

[21] C. Diaz, S. Seys, J. Claessens, and B. Preneel, "Towards Measuring Anonymity", presented at Privacy Enhancing Technologies (PET), 2002.

9. Appendix

Table 1. Notations and Terminology

DT_n	Latest time when the tracking system monitored node *n* with its old pseudonym
DP_n	Last location where the tracking system monitored node *n* with its old pseudonym
ET_n	Earliest time when the tracking system monitored node *n* with its new pseudonym
EP_n	First location where the tracking system monitored node *n* with its new pseudonym
C_n^m	Combination operator
GAS	Geographical anonymity set
TA	Tracking accuracy
Target	Node whose privacy is being measured
Mixer node	Nodes (excluding the target) that participate in the location privacy protection system
$PC_{x,y,t}$	Discrete random variable, which represents the privacy contribution at position (x,y) and time t, and whose range of values is $\{\alpha 1, \alpha 2, \ldots \alpha l, \ldots\}$
Q_t	Continuous random variable, which represents the target's destination distribution after t seconds.
α_i	Value of the privacy contribution, which depends only on the number of nodes arriving simultaneously with the target at same location
$p_{x,y,t}(i,j)$	Probability that for a given position (x,y) there will be i nodes at (x,y), and that the emerging times of j nodes from these i nodes overlap with that of target
$r_{x,y,t}(i)$	Probability that i nodes arrive at position (x,y) after t seconds
q(x,y,t,n)	Probability distribution function of Qt within two dimensional space for a given node n after t seconds
$PPC_n(x,y,t)$	Point privacy contribution at position (x,y) after t seconds
NPL(t,n)	Node privacy level of node n after t seconds
SPL(t)	System privacy level after t seconds
T1	Address lifetime
T2	Silent period, which varies between [*T2min*, *T2max*], with a range of $\triangle$T2.
b(i,j)	Probability density function of the emerging time of the j^{th} mixer (from the total i arrived mixers) that overlaps with that of the target