On the Age of Pseudonyms in Mobile Ad Hoc Networks

Julien Freudiger*, Mohammad Hossein Manshaei*, Jean-Yves Le Boudec, and Jean-Pierre Hubaux
Laboratory for computer Communications and Applications (LCA), EPFL, Switzerland
{julien.freudiger, hossein.manshaei, jean-yves.leboudec, jean-pierre.hubaux}@epfl.ch

***Abstract*—In many envisioned mobile ad hoc networks, nodes are expected to periodically beacon to advertise their presence. In this way, they can receive messages addressed to them or participate in routing operations. Yet, these beacons leak information about the nodes and thus hamper their privacy. A classic remedy consists in each node making use of (certified) pseudonyms and changing its pseudonym in specific locations called mix zones. Of course, privacy is then higher if the pseudonyms are short-lived (i.e., nodes have a short distance to confusion), but pseudonyms can be costly, as they are usually obtained from an external authority. In this paper, we provide a detailed analytical evaluation of the age of pseudonyms based on differential equations. We corroborate this model by a set of simulations. This paper thus provides a detailed quantitative framework for selecting the parameters of a pseudonym-based privacy system in peer-to-peer wireless networks.**

I. Introduction

Virtually all deployed wireless networks require mobile nodes to communicate in a single hop with the (wired) infrastructure, typically through a base station or an access point. However, the growing popularity of Bluetooth, WiFi in ad hoc mode, and other similar techniques are likely to fuel the adoption of peer-to-peer wireless communications. In that case, wireless nodes communicate directly with each other over a single hop or over multiple hops. This capability can be used to support a number of applications, ranging from urban sensing to mobile social networks to vehicular ad hoc networks (VANET). Of course, peer-to-peer wireless communications can coexist with the aforementioned classic wireless networks. In this paper, we focus exclusively on the former.

In most peer-to-peer wireless communication systems, each node is expected to periodically beacon to advertise its presence. In this way, it can receive messages addressed to it or participate in routing operations. Yet, these beacons leak information about the node and thus hamper its privacy. In particular, external parties can monitor beacons to learn the locations of mobile nodes.

A classic remedy to protect the location privacy of mobile nodes consists in relying on multiple pseudonyms: a node uses a pseudonym for a while, then discards it and makes use of a new one. This requires each node to have a repository of pseudonyms that it refills whenever needed. In many cases, these pseudonyms are used by other entities (e.g., other nodes) as trustworthy identifiers for authentication and thus need to be certified by a trusted certification authority. The pseudonym mechanism must thus be designed with great care, because information about the identity of the node can potentially be leaked at various protocol layers, notably by the IP and MAC address [13]. But even with these precautions, changing pseudonyms from time to time might not be enough, because the adversary can track mobile nodes spatially and temporally [2]. As a consequence, nodes should change their pseudonyms in a coordinated fashion with their neighbors in *mix zones* [4]. In other words, location privacy cannot be achieved by itself and requires a collective effort from neighboring mobile nodes.

The age of a pseudonym refers to the time period over which a given pseudonym is used. Of course, privacy is higher if the pseudonyms are short-lived. Yet, pseudonyms are costly, as they are usually obtained from an external authority and because a change of pseudonym is a burden for a node: that change can mean remaining unreachable for a short while (typically during the sojourn in the mix zone), entailing the loss of ongoing transactions, or requiring the update of routing tables. Consequently, in many cases a node might consider that its level of privacy is still high enough and might prefer to *not* change its pseudonym, even if it is located in a mix zone.

The coordination of pseudonym changes among nodes with different privacy levels is thus a central problem to achieve location privacy with multiple pseudonyms. Several approaches [4], [6], [11], [12], [14], [21] make use of an infrastructure or rely on pre-determined time/location to coordinate the pseudonym changes of mobile nodes. However, such solutions require the help of the infrastructure or that mobile nodes learn prior to entering the network the location of mix zones. Several researchers [10], [20], [21], [23] advocate the use of a distributed solution, where mobile nodes coordinate pseudonym changes to dynamically obtain mix zones. This solution is particularly appealing to mobile ad hoc networks because it does not require the help of the infrastructure.

In a distributed setting, it remains unclear how successful nodes will be in coordinating their pseudonym changes and how it will affect the age of their pseudonyms. Most existing evaluations do not model the dynamics of the system and consequently do not provide critical conditions for the success of the multiple pseudonym approach. In this paper, we push this distributed approach further and provide a framework for analytically evaluating the privacy obtained with mix zones. That framework captures the mobility of the nodes and the evolution of their privacy level over time. It provides designers

[*] Equally contributing authors.

with *conditions for the emergence of location privacy in mobile ad hoc networks*. We validate our analytical results with simulations.

The paper is organized in the following way. Section II contains a detailed model of the considered system. Section III is focused on the analytical model, expressed in differential equations, the solution of which is provided in Section IV. The simulations corroborating these results are provided in Section V. Finally, Section VI contains a discussion of the related work and Section VII concludes the paper.

II. System Model

In this section, we introduce the assumptions made throughout the paper.

A. Network Model

We study a network where mobile nodes are autonomous entities equipped with WiFi or Bluetooth-enabled devices that communicate with each other upon coming into radio range. In other words, we consider a mobile wireless system such as a vehicular network or a network of directly communicating hand-held devices. Without loss of generality, we assume that each user in the system has a single mobile device and thus corresponds to a single node in the network.

As commonly assumed in such networks, we consider an offline central authority (CA) run by an independent trusted third party that manages, among other things, the security and privacy of the network. In line with the multiple pseudonym approach, we assume that prior to joining the network, every mobile node in $\{u_i\}_{i=1}^{N}$, where N is the total number of mobile nodes in the system, registers with the CA that preloads a finite set of *pseudonyms* (e.g., certified public/private key pairs, MAC addresses). Mobile nodes change pseudonyms in mix zones in order to achieve location privacy. Upon changing pseudonyms, we consider for simplicity that the old pseudonym expires and is removed from the node's memory. Once a mobile node has used all its pseudonyms, it contacts the CA to obtain a new set of pseudonyms.

We assume that mobile nodes automatically exchange information (unbeknownst to their users) as soon as they are in communication range. Note that our evaluation is independent of the communication protocol. Without loss of generality, we assume that mobile nodes advertise their presence by periodically broadcasting proximity beacons containing the node's identifying information (i.e., the sender attaches its pseudonym to its messages). Due to the broadcast nature of wireless communications, beacons enable mobile nodes to discover their neighbors. For example, when a node receives an authenticated beacon, it controls the legitimacy of the sender by checking the certificate of the public key of the sender. After that, the node verifies the signature of the beacon message.

We consider a discrete time system with initial time $t = 0$. At each time step t, each mobile node can move independently of others on a plane in the considered area. We consider a random-trip mobility model characterized by the rate of encounters η, and the average number of nodes met in an encounter $\bar{N}$. The rate η determines the number of encounters with nearby nodes that occur on average. The average $\bar{N}$ establishes the average number of nodes that participate in each encounter. The meeting rate η and the average $\bar{N}$ depend on nodes' speed and the topology of the underlying road network. In our simulations, we consider a random walk model [7] satisfying predetermined η and $\bar{N}$ values.

B. Threat Model

An adversary $\mathcal{A}$ aims at tracking the location of some mobile nodes. In practice, the adversary can be a rogue individual, a set of malicious mobile nodes, or might even deploy its own infrastructure (e.g., by placing eavesdropping devices in a given area). We assume that the adversary is *passive* and simply eavesdrops on communications. In the worst case, $\mathcal{A}$ obtains complete coverage and tracks mobile nodes throughout the entire area. We characterize the latter type of adversary as *global*.

$\mathcal{A}$ collects identifying information (e.g., the MAC address or the public keys used to sign messages) from the entire network and obtains *location traces* that allow him to track the location of mobile nodes. The problem we tackle in this paper consists in protecting the *location privacy* of mobile nodes, that is, in preventing other parties from learning a node's past and current location [4]. It must be noted that, at the physical layer, the wireless transceiver has a wireless fingerprint that the adversary could use to identify it [5], [9], [15], [24]. However, this requires a costly installation for the adversary and stringent conditions on the wireless medium; in addition, countermeasures could be developed. Hence, it remains unclear how much identifying information can be extracted in practice from the physical layer and we do not consider this threat. Finally, note that higher layer defenses such as mix zones can be useful whether or not physical layer attacks are in place. For example, some applications may need to store location data to do congestion analysis.

C. Location Privacy Model

There are several techniques to mitigate the tracking of mobile nodes. In this paper, we consider the use of *multiple pseudonyms*: mobile nodes change over time their pseudonym to reduce their long term linkability.

1) Mix Zones: Mobile nodes in proximity of each other coordinate pseudonym changes in regions called mix zones in order to avoid temporal correlation of their locations. Mix zones can also conceal the trajectory of mobile nodes in order to protect against the spatial correlation of location traces, e.g., by using (i) silent mix zones [20], [23], (ii) a mobile proxy [25], (iii) regions where the adversary has no coverage [6], or (iv) encrypted communications [11]. Without loss of generality, we assume silent mix zones: mobile nodes turn off their transceivers and stop sending messages for a certain period of time. If at least two nodes change pseudonyms in a silent mix zone, a mixing of their whereabouts occurs and the mix zone becomes a *confusion point* for the adversary.

Consider a group of $n(t)$ mobile nodes at time t that are in proximity. One node among the $n(t)$ nodes can initiate a change of pseudonym using the one-round protocol suggested in [23] (i.e., the Swing protocol): a mobile node broadcasts an initiation message to start the pseudonym change. The $n(t)-1$ mobile nodes in proximity receive the message and enter a silent period during which they decide whether to change their pseudonyms or not. During the silent period, nodes do not communicate with each other. At the end of the silent period, it appears as if all pseudonym changes have occurred simultaneously. Mobile nodes must thus decide to change pseudonyms without knowing the decision of other nodes in proximity.

The adversary $\mathcal{A}$ observes the set of $n(T)$ nodes changing pseudonyms, where T is the time at which the pseudonym change occurs. $\mathcal{A}$ compares the set B of pseudonyms before the change with the set R of pseudonyms after the change and, based on the mobility of the nodes, predicts the most probable matching [4], [23]. Let $p_{r|b} = Pr$("Pseudonym $r \in R$ corresponds to $b \in B$"), that is the probability that a new pseudonym $r \in R$ corresponds to an old pseudonym $b \in B$. As usually done in the literature [26], the uncertainty of the adversary (or the untraceability of node u_i using pseudonym b) is defined as:

$$A_i(T) = -\sum_{r=1}^{n(T)} p_{r|b} \log_2(p_{r|b}) \tag{1}$$

The achievable location privacy depends on both the number of nodes $n(T)$ and the mobility of the nodes $p_{r|b}$ in the mix zone. If a node u_i changes its pseudonym alone, then the adversary can track node u_i, and we write $A_i(T) = 0$. The entropy is maximum for a uniform probability distribution $p_{r|b}$ and the achievable location privacy after a coordinated pseudonym change at time T is upperbounded by $\log_2(n(T))$. If at least two mobile nodes (including u_i) change their pseudonyms, then the pseudonym change is successful and generates a confusion point. We denote T_i^ℓ the time of the *last* successful pseudonym change of node u_i.

2) Distance to Confusion or the Age of Pseudonyms: As observed in [19], the degree of location privacy not only depends on the location privacy achieved in mix zones by the nodes traversing it, but also on how long an adversary can successfully track mobile nodes between mix zones. A longer tracking period increases the likelihood that the adversary identifies the mobile nodes. Hence, mobile nodes should evaluate the distance over which they are potentially tracked by an adversary (i.e., the *distance to confusion* [17]) and act upon it by deciding to change pseudonyms accordingly. To capture the notion of distance to confusion, we define the *age of a pseudonym* as the time period over which a given pseudonym is used.

In this work, we model the evolution of the age of pseudonyms over time $Z_i(t)$ for each mobile node u_i as a linearly increasing function of time with an *aging rate* λ_i:

$$Z_i(t) = \lambda_i \cdot (t - T_i^\ell) \tag{2}$$

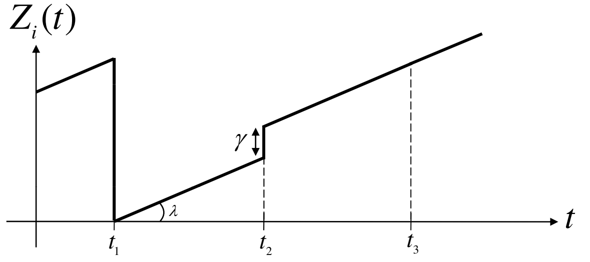


Fig. 1. Example of evolution of the age of pseudonyms. At t_1, node u_i successfully changes pseudonym with another node and the age of its pseudonym drops to zero. The age of pseudonym of node u_i then increases with rate λ. At t_2, the pseudonym change fails and node u_i pays the cost γ of changing a pseudonym. At t_3, the node refuses to change its pseudonym.

where t is the current time and $T_i^\ell \leq t$ is the time of the last successful pseudonym change of mobile u_i. The value $Z_i(t)$ captures the age of the current pseudonym of user i at time t (Fig. 1). The aging rate λ_i mainly depends on the belief of node u_i with respect to the tracking power of the adversary and on the beaconing rate/range of node u_i. The higher the value of λ_i is, the faster the pseudonyms age. For simplicity, we consider that $\lambda_i = \lambda$, $\forall i$.

3) Strategies: With this model, mobile nodes request a pseudonym change when the age of their pseudonym is considered large and if there are other nodes in proximity. Nodes in proximity choose to cooperate (C) or defect (D) if their pseudonym age is large as well. Hence, the success of a pseudonym change depends on the state of the neighboring nodes. Asynchronous requests to change pseudonyms might cause failed attempts to achieve location privacy. Assume that n_c is the number of nodes that cooperate (change pseudonyms) in a meeting besides u_i and that $Z_i(t^-)$ is the age of u_i just before making its decision. Considering an encounter in a mix zone at time t, we write for node u_i:
If $C \wedge (n_c > 0)$,

$$T_i^\ell = t \tag{3}$$

$$Z_i(t) = 0 \tag{4}$$

If $C \wedge (n_c = 0)$,

$$Z_i(t) = Z_i(t^-) + \gamma \tag{5}$$

If D,

$$Z_i(t) = Z_i(t^-) \tag{6}$$

In other words, $Z_i(t)$ is reset to 0 when a pseudonym change is successful. If a node is alone in changing its pseudonym, then it pays the cost of changing pseudonym γ and in addition, the age of its pseudonym keeps increasing. The cost γ can be expressed as: $\gamma = \gamma_{acq} + \gamma_{rte} + \gamma_{sil}$, where γ_{acq} is the cost of acquiring new pseudonyms, γ_{rte} is the cost of updating routing tables, and γ_{sil} is the cost of remaining silent while traversing the mix zone. The cost γ is expressed in age units (i.e., time), causing an increase in the age of pseudonyms. The cost γ captures the failed opportunity of a pseudonym change and

TABLE I
LIST OF SYMBOLS.

Symbol	Definition
t	Time
λ	Pseudonym aging rate
γ	Cost of changing pseudonym
N	Total number of nodes in the system
$f(z,t)$	Probability distribution function of age of pseudonyms at time t
$F(z,t)$	Cumulative distribution function of age of pseudonyms at time t
$c(z)$	Probability of cooperation of mobile nodes with age z
$q(t)$	Probability that at least one node in a meeting at time t cooperates
η	Rate of meetings
$\bar{c}(t)$	Probability of cooperation for a randomly selected node at time t
h_n	Probability that a meeting involves $n+1$ nodes
$H(\mathbf{z})$	$\mathcal{Z}$-transform of h_n
T_i^ℓ	Time of last successful pseudonym change for node u_i
$Z_i(t)$	Age of pseudonym of node u_i at time t
$\bar{N}$	Average number of nodes in a meeting
z	Age of pseudonym
θ	Threshold for cooperation

is thus an incentive to carefully manage pseudonyms. Finally, if a node defects, its pseudonym age is unchanged. Figure 1 illustrates how the age of pseudonyms evolves with time in the case of meetings between several nodes. With this model, nodes control the distance over which they can be tracked.

Mobile nodes decide when to change pseudonyms next based on the time of their last successful pseudonym change T_i^ℓ. We define $c_i(z)$ the probability distribution over the age Z_i that gives the probability of cooperation of each node u_i. For simplicity, we assume that the distribution is the same for all nodes and we write $c_i(z) = c(z)$. Hence, when several nodes meet, each node decides whether to change its pseudonym with probability $c(z)$.

D. Metric

We are interested in measuring the success of the multiple pseudonym approach. Changing a pseudonym is successful only if it is coordinated with other nodes nearby. Hence, in order to evaluate the ability of nodes to synchronize, we measure the distribution of the age of pseudonyms in the network. We define $Z(t) \sim f(z,t)$ a random variable that describes the density of probability for any age z. The CDF $F(z,t) = \int_z f(x,t)dx$ gives the fraction of nodes u_i at time t whose age of pseudonym is $Z_i(t) \leq z$.

III. Analytical Evaluation

In this section, we derive analytically the probability distribution of the age of pseudonyms, $F(z,t)$. To do so, we calculate the fraction of users whose age of pseudonyms is lower than z, i.e. $Pr\{Z \leqslant z\}$. We show that the evolution of the age of pseudonyms can be approximated by a dynamical system composed of a simple differential equation when the number of nodes N gets large. Table I summarizes the notation used throughout the paper.

A. Dynamical System

As discussed above, the random variable $Z(t)$ models the distribution of the age of pseudonyms at time t. The evolution of this random variable over time can be captured by a dynamical system composed of *drift* and *jump processes*. The goal of the *drift* and *jump* processes is to capture the dynamics of the age of pseudonyms by modeling the possible variations of a pseudonym age. The drift models an increase in the age of the pseudonyms (e.g., aging of a pseudonym), and a jump models sudden changes in the age of the pseudonyms (e.g., upon using a new pseudonym).

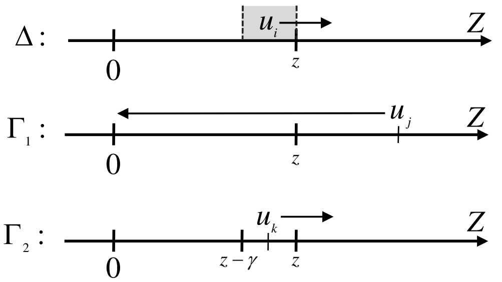


Fig. 2. Illustration of the drift (Δ) and jump (Γ_1 and Γ_2) processes. We observe the three scenarios causing a change in the distribution $F(z,t)$.

1) Drift Process Δ*:* The drift process models the aging of pseudonyms over time as shown in Fig. 2. At each time step, the age of the pseudonym of every node is incremented with rate λ. Hence, for any fixed value z, the age of the pseudonyms of a fraction of nodes will pass above z and decrease $F(z,t)$. The drift directly depends on the aging rate λ and the density of the age of the pseudonyms:

$$\Delta = \lambda \frac{\partial F}{\partial z} \tag{7}$$

2) Jump Process Γ*:* The jump process captures the sudden variations in the age of pseudonyms. There are two possible scenarios, Γ_1 and Γ_2, that correspond to successful and failed attempts to change pseudonyms, respectively (Fig. 2). In the first type of jump Γ_1, a node u_j with a pseudonym age larger than z successfully changes its pseudonym with other encountered nodes in proximity. Hence, the age of its pseudonym drops to 0. This happens with rate:

$$\Gamma_1 = \eta \int_0^\infty c(x)q(t)(1 - 1_{\{x \leq z\}})\frac{\partial F}{\partial x}(x,t)dx \tag{8}$$

where $q(t)$ is the probability that at least one of the encountered nodes changes pseudonym as well; $c(z)$ is the probability of cooperation of user u_j given that its age of pseudonym is z, and η is the rate of meetings scaled by the number of nodes. Intuitively, Γ_1 (i.e., the rate at which any node u_j successfully changes pseudonym) depends on the rate of encounter between nodes η, on the probability that u_j cooperates $c(z)$, on the probability of meeting a nodes that cooperates $q(t)$, and on the probability of having a pseudonym age larger than z. The integral captures the probability that a node u_j has a pseudonym age larger than z, cooperates, and meets at least one cooperative node, thus causing an increase in $F(z)$.

In the second type of jump process Γ_2, a user u_k with a pseudonym age between $z - \gamma$ and z changes pseudonym in an encounter with other nodes. However, none of the nodes in proximity cooperate, and the pseudonym change is a failure. Hence a pseudonym is wasted and user u_k suffers a cost γ: $Z_k(t) = Z_k^- + \gamma$ causing an increase in the number of users with an age of pseudonym larger than z. This happens with rate:

$$\Gamma_2 = \eta \int_{z-\gamma}^{z} c(x)(1-q(t))\frac{\partial F}{\partial x}(x,t)dx \tag{9}$$

Intuitively, Γ_2 (i.e., the rate at which u_k fails to change pseudonyms) depends on the rate of encounter η, on the probability that u_k cooperates, on the probability of meeting nodes that all defect $(1-q(t))$, and on the probability that u_k has a pseudonym age in the interval $[z-\gamma, z]$. The integral captures the probability that a node u_k has a pseudonym age in the interval $[z-\gamma, z]$, cooperates, and meets nodes that all defect, thus causing a decrease in $F(z)$.

B. Differential Equation

Taking into account the drift and jump processes, we obtain a dynamical system defined by a single differential equation. The cumulative distribution function $F(z,t)$, giving the fraction of nodes with an age of pseudonym smaller than z, is the unique solution of the following differential equation:

$$\frac{\partial F}{\partial t} = -\Delta + \Gamma_1 - \Gamma_2 \tag{10}$$

with boundary conditions: $F(\infty, t) = 1, \forall t$. Intuitively, on one hand, the drift Δ and the jump Γ_2 cause nodes to have an age larger than z, hence decreasing the fraction of nodes $F(z)$. For this reason, they are subtracted from $\frac{\partial F}{\partial t}$. On the other hand, the jump Γ_1 increases the number of nodes on the left size of z, hence increasing $F(z)$. For this reason, it is added to $\frac{\partial F}{\partial t}$.

As defined above, $q(t)$ is the probability that at least one of the encountered nodes cooperates. It can be calculated by considering the probability of meeting n nodes and the probability that at least one node cooperates:

$$q(t) = 1 - \sum_{n\geq 0} h_n(1-\bar{c}(t))^n = 1 - H(1-\bar{c}(t)) \tag{11}$$

where h_n is the probability of meeting n nodes (a meeting involves $n+1$ nodes: the node itself with the n encountered nodes), $H(\mathbf{z}) = \sum_{n\geqslant 0} \mathbf{z}^n h_n$ is the $\mathcal{Z}$-transform of h_n, and $\bar{c}(t)$ is the probability that an encountered node cooperates:

$$\bar{c}(t) = \int_0^\infty c(z)f(z,t)dz \tag{12}$$

Intuition of the equation above: The main idea of our approach is to replace all interactions between nodes with an average interaction. This can be done by using the principles of Mean Field theory. To do so, we consider the probability that each node has a certain age in the system (e.g., $f(z)$). Previous work [1], [8] has shown that such probability distribution function converges to a deterministic limit (mean field convergence) when N goes to infinity. The probability distribution function is known to satisfy an ordinary differential equation formed by drift and jump processes that capture the possible transitions in the age of pseudonyms. In summary, by considering the possible scenarios that affect the age of pseudonyms, we derive the above differential equation characterizing the distribution of the age of pseudonyms.

IV. Analytical Results

In this section, we solve the differential equation (10) characterizing the age of pseudonyms. We consider the system in the stationary regime (i.e., as t goes to infinity, we have $\frac{\partial F}{\partial t} = 0$) and evaluate how system parameters such as η, λ, θ, and $c(z)$ affect the distribution of the age of pseudonyms $F(z,t)$.

We assume that a node cooperates according to a simple threshold function. This means that if its age of pseudonym is smaller than a given threshold θ, it decides not to cooperate, whereas it cooperates with probability c_0 if its age of pseudonym is larger than θ:

$$c(z) = \begin{cases} 0 & z \leqslant \theta \\ c_0 & z > \theta \end{cases} \tag{13}$$

Intuitively, a node will tend not to cooperate as long as it estimates that the age of its pseudonym (or its distance to confusion) is sufficient. With this model, the threshold θ and the probability c_0 determine the inclination of each node to cooperate. For example, a low θ and a high c_0 mean that the nodes will often change their pseudonyms. These parameters directly affect the probability distribution of the age of pseudonyms. Consequently, we can fine tune the achievable level of privacy in the system.

As mentioned, we have $\frac{\partial F}{\partial t} = 0$ in the stationary regime. For simplicity, we derive Equation (10) with respect to z and as $\frac{\partial F}{\partial z}(z,t) = f(z,t)$, we obtain:

$$\begin{cases} \lambda \frac{\partial f}{\partial z} + \eta c(z) f(z) - \eta(1-q)c(z-\gamma)f(z-\gamma) = 0 \\ \int_0^\infty f(z)dz = 1 \end{cases} \tag{14}$$

Considering the probability of cooperation $c(z)$ defined by Equation (13), the above differential equation must be solved in three intervals:

1) $z < \theta$: The probability of cooperation $c(z)$ is equal to 0 in this interval (i.e., nodes never cooperate). Hence the differential equation (14) becomes $\frac{\partial f}{\partial z} = 0$. The solution is then $f(z) = f(0)$.
2) $\theta \leqslant z < \theta + \gamma$: The probability of cooperation $c(z-\gamma)$ is equal to 0 in this interval and the differential equation is: $\lambda \frac{\partial f}{\partial z} + \eta c_0 f(z) = 0$. Considering the boundary condition $f(z = \theta)$, the solution in this interval is: $f(z) = f(0)e^{\frac{-\eta c_0}{\lambda}(z-\theta)}$.
3) $\theta + \gamma \leqslant z$: For these values of z, the differential equation (14) is a non-autonomous differential equation. We iteratively solve this differential equation by solving a series of autonomous differential equations in the interval $[0, \gamma]$. As illustrated in Fig. 3, we define m functions f_m, $m = 1, 2, 3, \cdots, \infty$, over the interval $[0, \gamma]$. For each

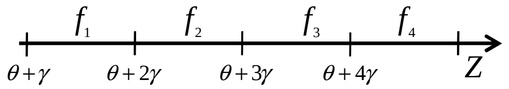


Fig. 3. The definition of f_m over $[\theta + m\gamma, \theta + (m + 1)\gamma]$ intervals.

interval, we obtain an autonomous differential equation as follows:

$$\begin{cases} \frac{\partial f_m(x)}{\partial x} + \frac{\eta c_0}{\lambda} f_m(x) - \frac{\eta c_0(1-q)}{\lambda} f_{m-1}(x) = 0 \\ f_m(0) = f_{m-1}(\gamma) \qquad m = 1, 2, 3, \cdots, \infty \end{cases} \tag{15}$$

In order to ensure the continuity of $f(z)$, we need to take into account the solution of $f(z)$ in the interval $[\theta, \theta + \gamma]$. Hence, we know that $f_0(\gamma)$ must be equal to $f(0)e^{\frac{-\eta c_0}{\lambda}(\gamma)}$. For $m = 1, 2, 3, \cdots, \infty$, the solution of the above system of iterative equations is:

$$f_m(x) = e^{\frac{-\eta c_0 x}{\lambda}} \left(f_{m-1}(\gamma) + \frac{\eta c_0(1-q)}{\lambda} \int_0^x e^{\frac{\eta c_0 y}{\lambda}} f_{m-1}(y) dy \right) \tag{16}$$

Finally, we obtain the values of $f(z)$ by calculating $f_m(z-(\theta+m\gamma))$ for every interval $[\theta+m\gamma, \theta+(m+1)\gamma]$.

Let us define $\alpha = \eta c_0/\lambda$ for simplicity. After some simplifications on Equation (16), we obtain:

$$f(z) = \begin{cases} f(0) & z < \theta \\ f(0)e^{-\alpha(z-\theta)} & \theta \leqslant z < \theta + \gamma \\ f(0)e^{-\alpha(z-\theta)} g(z) & \theta + \gamma \leqslant z \end{cases} \tag{17}$$

where $g(z)$ is a polynomial function as follows:

$$g(z) = \sum_{k=0}^{m} \frac{\alpha^k}{k!} e^{k\alpha\gamma} (1-q)^k (z - k\gamma - \theta)^k \tag{18}$$

for $\theta + m\gamma \leq z < \theta + (m + 1)\gamma$. Recall that $f(0)$ can be calculated using the boundary condition presented in Equation (14). After some simplification, we obtain:

$$f(0) = \frac{1}{\theta + \frac{1-e^{-\alpha\gamma}}{\alpha} + I} \tag{19}$$

where $I = \sum_{m=1}^{\infty} \sum_{k=0}^{m} e^{\alpha(\theta+k\gamma)} \frac{(1-q)^k \alpha^k}{k!} \int_{\theta+m\gamma}^{\theta+(m+1)\gamma} e^{-\alpha z}(z - \theta - k\gamma)^k dz$.

From the above equation, we observe that we need to calculate probability q (at least one node cooperates at the meeting point) in order to obtain $f(z)$. To do so, we must compute probability h_n of meeting n nodes and the probability of cooperation of a node $\bar{c}$ as shown in Equation (11).

A. Derivation of Probability q

Assume that the average number of nodes in a meeting point is $\bar{N}$. Usually, the probability of having n nodes in a meeting point follows a long tail distribution. In this paper, we consider a Geometric distribution with parameter w for the probability of meeting n nodes. By definition of a Geometric definition, the average number of nodes at a meeting point is $\bar{N} = \frac{w}{1-w}$. The probability of meeting n nodes is then:

$$h_n = w^n(1 - w) \tag{20}$$

Hence, the $\mathcal{Z}$-transform of h_n is:

$$H(\mathbf{z}) = \sum_{n \geqslant 0} \mathbf{z}^n w^n (1 - w) = \frac{\mathbf{z}(1 - w)}{1 - \mathbf{z}w} \tag{21}$$

We also need to compute the average probability of cooperation $\bar{c}$. Using Equation (12), we obtain:

$$\begin{aligned} \bar{c} &= f(0)c_0 \left(\frac{1 - e^{-\alpha\gamma}}{\alpha} + \int_{\theta+\gamma}^{\infty} e^{-\alpha(z-\theta)} g(y) dy \right) \\ &= f(0)c_0 \left(\frac{1 - e^{-\alpha\gamma}}{\alpha} + I \right) \end{aligned} \tag{22}$$

Finally, q is obtained by computing $H(1 - \bar{c})$, which is the value between 0 and 1 that satisfies the following equation:

$$\frac{c_0}{q} = \frac{1}{w} - (1 - c_0) + \frac{\theta(1 - w)}{w(\frac{1-e^{-\alpha\gamma}}{\alpha} + I)} \tag{23}$$

Our results in Equation (17) show that the probability density function $f(z)$ will first be uniform in the interval $[0, \theta]$. Then, on the small interval $[\theta, \theta + \gamma]$, it will decrease exponentially. For the other values of z, the probability $f(z)$ decreases according to an exponential distribution multiplied by a polynomial $g(z)$. Intuitively, it means that nodes will be evenly distributed below the threshold θ and for the other values of $z > \theta$ will have a long tail distribution.

With respect to probability q, we observe that it not only depends on cooperation parameters such as c_0 and θ, but also depends on the rate of encounters η, the average number of nodes in an encounter $\bar{N}$, and on the cost of changing pseudonym γ.

B. Example with $\gamma = 0$ and $c_0 = 1$

Assume that the cost of changing pseudonym $\gamma = 0$ and that $c_0 = 1$. The probability distribution function $f(z)$ can be rewritten as

$$f(z) = \begin{cases} f(0) & z < \theta \\ f(0)e^{\frac{-\eta q}{\lambda}(z-\theta)} & \theta \leqslant z \end{cases} \tag{24}$$

where $f(0) = \frac{1}{\theta + \frac{\lambda}{\eta q}}$. We compute q by using Equation (11). Considering our threshold cooperation function $c(z)$, $\bar{c}(t)$ is:

$$\bar{c} = \int_{\theta}^{\infty} f(0)e^{-\frac{\eta q}{\lambda}(z-\theta)} dz = \frac{\lambda}{\lambda + \eta q \theta} \tag{25}$$

Finally q can be calculated by replacing $\bar{c}$ in Equation (11):

$$q = \frac{\lambda - \sqrt{\lambda(4\eta\theta w(1 - w) + \lambda)}}{2(-1 + w)\eta\theta} \tag{26}$$

We observe that in this simple example, the probability density function $f(z)$ is first uniform for $z < \theta$ and then decreases according to an exponential distribution.

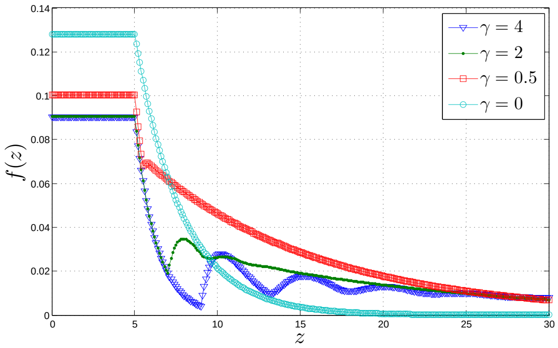


Fig. 4. Probability distribution function $f(z)$ for different values of γ.

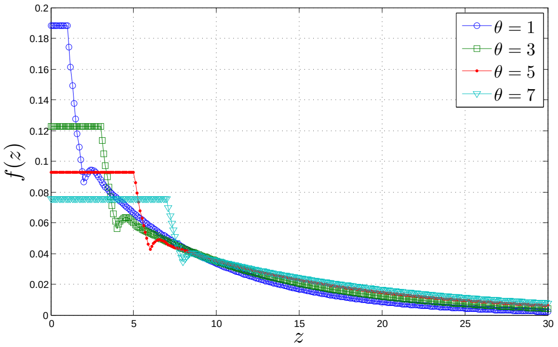


Fig. 5. Probability distribution function $f(z)$ for different values of θ.

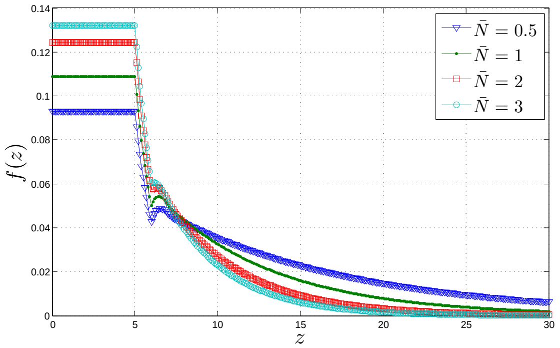


Fig. 6. Probability distribution function $f(z)$ for different values of $\bar{N}$.

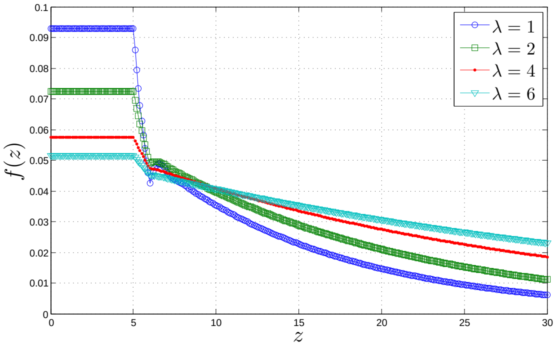


Fig. 7. Probability distribution function $f(z)$ for different values of λ.

C. Numerical Evaluation

In this section, we evaluate numerically the analytical results of the previous section. In particular, we study how the system parameters affect the distribution of the age of pseudonyms, $f(z)$. Unless otherwise stated, we use the following values for the system parameters: $\bar{N} = 0.5$, $\theta = 5$, $\gamma = 1$, $\lambda = 1$, $\eta = 0.75$, and $c_0 = 1$.

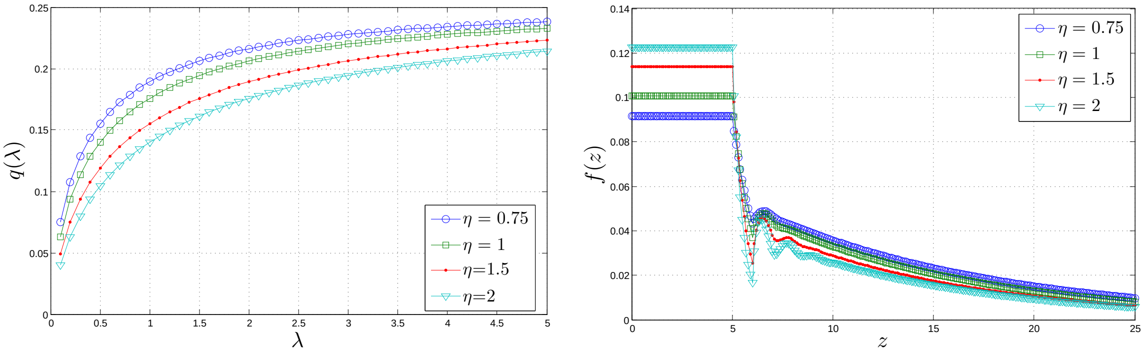
As shown in Equation (17), the probability density function $f(z)$ has three different behaviors: it is first constant with value $f(0)$, then decreases exponentially with parameter $-\alpha$ and finally decreases according to an exponential multiplied by a polynomial which is different for every interval of size γ. We observe in Fig. 4 the three behaviors of $f(z)$ for different values of γ. For example, with $\gamma = 4$, we have $f(z) = f(0) = 0.09$ over the interval $[0, \theta]$. Then, $f(z)$ exponentially decreases until $\theta + \gamma = 9$. Finally, we observe that $f(z)$ oscillates because of the polynomial function (18) which is different for every interval of size γ. As z increases, the oscillation is attenuated because the exponential term dominates the polynomial function. Intuitively, the oscillation is caused by the jump process Γ_2: nodes with age of pseudonym belonging to $[z - \gamma, z]$ fail to coordinate and their age of pseudonym is thus increased by γ. In Fig. 4, we also observe the effect of different values of γ on the distribution $f(z)$. As γ decreases, the oscillations become less noticeable because the jump process Γ_2 affects fewer nodes (since the interval $[z - \gamma, z]$ becomes smaller). Moreover, in the case of $\gamma = 0$, we notice that there is no oscillation because Γ_2 does not affect any node. Note that when γ decreases, more nodes have an age of pseudonym smaller than the threshold θ.

Figure 5 shows the effect of different θ on $f(z)$. We observe that with larger values of θ, the number of nodes with age of pseudonym below θ increases. A system designer can thus fine tune θ to vary the population of nodes with age of pseudonym smaller than θ. As θ increases, we notice that the average value of z increases as well, meaning that more nodes have a high age of pseudonym because nodes are less cooperative.

Figure 6 illustrates the effect of the average number of nodes $\bar{N}$ in meetings on $f(z)$. When $\bar{N}$ increases, the probability q to find a cooperative node increases and consequently the number of nodes with age of pseudonym below the threshold θ increases as well, meaning that in average the age of pseudonym is smaller.

Figure 7 illustrates the effect of the aging rate λ on $f(z)$. We observe that with a high λ (i.e., pseudonyms age faster), fewer nodes have an age of pseudonym below θ compared to lower values of λ.

Finally, we evaluate the influence of the rate of meetings η on $f(z)$ in Fig. 8. First, we focus on the probability q of encountering at least one cooperative node in Fig. 8 (a). As λ increases, nodes age faster and we observe that their probability of cooperation increases logarithmically. When the rate η increases, we observe that the probability of cooperation



(a)

(b)

Fig. 8. Influence of the rate of meetings η: (a) Probability q that at least one node in a meeting cooperates. (b) Probability distribution function $f(z)$.

q decreases for any value of λ: the reason is that for larger values of η, both jump processes Γ_1 and Γ_2 occur more frequently. Because Γ_1 dominates Γ_2 (as it affects more nodes), a larger fraction of nodes will have an age of pseudonym below θ (Fig. 8 (b)). For this reason, for a high η, fewer nodes cooperate, and q decreases.

The above results can help a system designer find the conditions for the emergence of location privacy. More specifically, the system designer can fine tune parameters such as θ, γ and λ in order to control the number of nodes with large age of pseudonym.

V. Validation with Simulations

In order to verify the relevance of our model, we compare our numerical evaluations with simulation results.

A. Simulation Setup

We consider a set of N mobile nodes moving according to a random walk model. The plane is composed of a grid of 10km×10km, where each step is of one meter. At every intersection, mobile nodes move from their current location to a new location by randomly choosing a direction. We consider that mobile nodes move with a constant speed. Directions are chosen out of $[0, 2\pi]$ with granularity $\pi/2$.

We consider that nodes are neighbors (i.e., in communication range) if they are within a fixed perimeter. We consider a communication range of 100m. Whenever a node has at least one neighbor, it must decide whether to cooperate or defect based on its value $Z_i(t)$ and the threshold cooperation function $c(z)$. After each iteration of the simulation, we compare the average of the current probability density function $f(z)$ to the average of $f(z)$ obtained in the 50 previous iterations. The simulation stops if the difference is smaller than 0.005, and otherwise runs at least for 200 iterations.

B. Simulation Results

Figure 9 compares $f(z)$ obtained with the numerical evaluation to the one obtained by simulation with two different values of $\gamma = 0$ and $\gamma = 4$. We consider the same value of η and $\bar{N}$ for the analytical and simulation results. The distribution of age obtained from the model shows a pretty good match with the distribution obtained with simulations. This means that our modeling assumptions succeeded in capturing the collective behavior of nodes changing their pseudonyms in mobile networks.

VI. Related Work

Previous works on location privacy [3], [18], [22] show that an adversary can implicitly obtain the true identity of the owner of a mobile node from the analysis of pseudonymous location traces. For example, using location traces collected in an office environment, Beresford and Stajano [3] correctly identified all participants by simply examining where the participants spent most of their time. Similarly, using GPS traces from vehicles, two studies by Hoh *et al.* [18] and Krumm [22] found the home (and thus the identity) of most drivers. Hence, pseudonyms are not sufficient to protect the location privacy of mobile nodes and should be changed *over time* to avoid such attacks. But even if location traces of mobile nodes do not contain any pseudonyms, Hoh and Gruteser [16] were able to reconstruct the tracks of mobile nodes using a multiple target tracking algorithm. Hence, location traces should also be altered *spatially*. In other words, the spatial and temporal correlation between successive locations of mobile nodes must be carefully eliminated to prevent external parties from compromising their location privacy. In this paper, location privacy is achieved by changing pseudonyms in regions called *mix zones* [3].

The coordination of pseudonym changes is thus a central problem to achieve location privacy with multiple pseudonyms and various solutions were proposed. One solution [6] consists in changing pseudonyms at a pre-determined frequency. The mechanism works if at least two mobile nodes change their pseudonyms in proximity. Other solutions suggest to use base stations as coordinators [21], or to change pseudonyms as specific time instance [14]. Another approach [4], [11], [12] coordinates pseudonym changes by forcing mobile nodes to change their pseudonyms within pre-determined regions called *mix zones*. Several researchers [10], [20], [21], [23] advocated the use of a distributed solution, where mobile nodes coordinate pseudonym changes to dynamically obtain

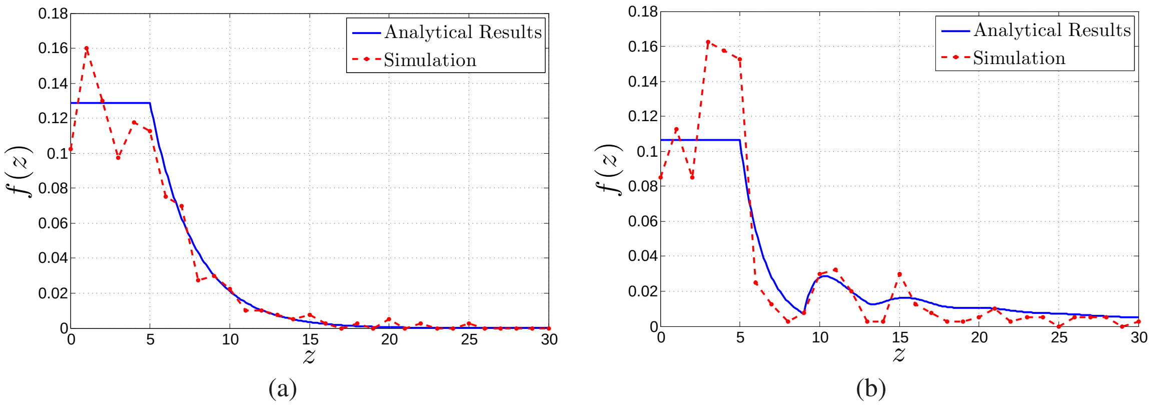


Fig. 9. Validation of numerical results: (a) $\gamma = 0$. (b) $\gamma = 4$.

mix zones. This solution is particularly interesting for mobile ad hoc networks because it does not require the help of the infrastructure, nor prior knowledge of the location of mix zones. In [10], nodes independently decide whether to change pseudonyms in mix zones. The decisions of the nodes is modeled with game theory and the authors show that selfish behavior dramatically decreases the chances of a successful coordination. However, none of the previous work measures the coordination success of pseudonym changes and its effect on the age of pseudonym.

VII. Conclusion and Future Work

We have considered the problem of achieving location privacy in mobile networks using multiple pseudonyms. We developed a framework to analytically evaluate the age of pseudonyms. Our framework captures the mobility and interactions between nodes. With this model, we obtained critical conditions for the emergence of location privacy. In particular, we evaluated the importance of the probability of cooperation of the nodes (θ and c_0), their mobility (η and $\bar{N}$), the cost of pseudonyms γ, and the aging rate λ. The model matches well with simulations, meaning that our modeling assumptions succeeded in capturing the collective behavior of nodes changing their pseudonyms in mobile networks. This paper is a first step towards obtaining a deeper understanding of the dynamics of privacy-preserving mechanisms in mobile networks.

In the future, we plan to consider other probability functions $c(z)$ for the cooperation of nodes. It would also be interesting to measure the relation between the level of uncertainty of an adversary and a given distribution of age of pseudonyms.

References

[1] M. Benaïm and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, 2008.

[2] A. R. Beresford. *Location Privacy in Ubiquitous Computing*. PhD thesis, University of Cambridge, 2005.

[3] A. R. Beresford and F. Stajano. Location privacy in pervasive computing. *Pervasive Computing, IEEE*, 2(1):46–55, 2003.

[4] A. R. Beresford and F. Stajano. Mix zones: user privacy in location-aware services. In *Pervasive Computing and Communications Workshops*, pages 127–131, 2004.

[5] V. Brik, S. Banerjee, M. Gruteser, and S. Oh. Wireless device identification with radiometric signatures. In *MobiCom*, 2008.

[6] L. Buttyan, T. Holczer, and I. Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs. In *ESAS*, 2007.

[7] T. Camp, J. Boleng, and V. Davies. A survey of mobility models for ad hoc network research. *Wireless Communication & Mobile Computing (WCMC)*, 2(5):483–502, 2002.

[8] A. Chaintreau, J.-Y. Le Boudec, and N. Ristanovic. The age of gossip: Spatial mean-field regime. In *ACM Sigmetrics*, 2009.

[9] B. Danev and S. Capkun. Transient-based identification of wireless sensor nodes. In *IPSN*, 2009.

[10] J. Freudiger, M. H. Manshaei, J.-P. Hubaux, and D. Parkes. On non-cooperative location privacy: A game-theoretic analysis. In *CCS*, 2009.

[11] J. Freudiger, M. Raya, M. Felegyhazi, P. Papadimitratos, and J.-P. Hubaux. Mix zones for location privacy in vehicular networks. In *WiN-ITS*, 2007.

[12] J. Freudiger, R. Shokri, and J.-P. Hubaux. On the optimal placement of mix zones. In *PETS*, 2009.

[13] B. Greenstein, D. McCoy, J. Pang, T. Kohno, S. Seshan, and D. Wetherall. Improving wireless privacy with an identifier-free link layer protocol. In *MobiSys*, 2008.

[14] M. Gruteser and D. Grunwald. Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis. *Mob. Netw. Appl.*, 2005.

[15] J. Hall, M. Barbeau, and E. Kranakis. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting. In *CIIT*, 2004.

[16] B. Hoh and M. Gruteser. Protecting location privacy through path confusion. In *SECURECOMM*, pages 194–205, 2005.

[17] B. Hoh, M. Gruteser, R. Herring, J. Ban, D. Work, J.-C. Herrera, A. M. Bayen, M. Annavaram, and Q. Jacobson. Virtual trip lines for distributed privacy-preserving traffic monitoring. In *MobiSys*, pages 15–28, 2008.

[18] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Enhancing security and privacy in traffic-monitoring systems. *IEEE Pervasive Computing*, 5(4):38–46, 2006.

[19] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady. Preserving privacy in GPS traces via path cloaking. In *CCS*, 2007.

[20] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki. Enhancing wireless location privacy using silent period. In *ECNC*, 2005.

[21] L. Huang, K. Matsuura, H. Yamane, and K. Sezako. Towards modeling wireless location privacy. In *PET*, 2005.

[22] J. Krumm. Inference attacks on location tracks. In *Pervasive*, 2007.

[23] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran. Swing & swap: User centric approaches towards maximizing location privacy. In *WPES*, 2006.

[24] B. Rasmussen and S. Capkun. Implications of radio fingerprinting on the security of sensor networks. In *SECURECOMM*, 2007.

[25] K. Sampigethaya, M. Li L. Huang, R. Poovendran, K. Matsuura, and K. Sezaki. CARAVAN: Providing location privacy for VANET. In *ESCAR*, 2005.

[26] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In *PET*, 2002.