Game-Theoretic Model of Incentivizing Privacy-Aware Users to Consent to Location Tracking

Emmanouil Panaousis[*], Aron Laszka[†], Johannes Pohl[‡],
Andreas Noack[‡], and Tansu Alpcan[§]
[*] University of Brighton, UK
[†] Institute for Software Integrated Systems, Vanderbilt University, Nashville, USA
[‡] University of Applied Sciences Stralsund, Germany
[§] University of Melbourne, Australia

***Abstract*—Nowadays, mobile users have a vast number of applications and services at their disposal. Each of these might impose some privacy threats on users' "Personally Identifiable Information" (PII). Location privacy is a crucial part of PII, and as such, privacy-aware users wish to maximize it. This privacy can be, for instance, threatened by a company, which collects users' traces and shares them with third parties. To maximize their location privacy, users can decide to get offline so that the company cannot localize their devices. The longer a user stays connected to a network, the more services he might receive, but his location privacy decreases. In this paper, we analyze the trade-off between location privacy, the level of services that a user experiences, and the profit of the company. To this end, we formulate a Stackelberg Bayesian game between the User (follower) and the Company (leader). We present theoretical results characterizing the equilibria of the game. To the best of our knowledge, our work is the first to model the economically rational decision-making of the service provider (i.e., the Company) in conjunction with the rational decision-making of users who wish to protect their location privacy. To evaluate the performance of our approach, we have used real-data from a testbed, and we have also shown that the game-theoretic strategy of the Company outperforms non-strategic methods. Finally, we have considered different User privacy types, and have determined the service level that incentivizes the User to stay connected as long as possible.**

***Index Terms*—Game theory, localization, privacy.**

I. Introduction

The prevalence of smartphones brings to end users not only new applications and services but also privacy risks. These risks are due to the possible disclosure of vast amount of private information. In this paper, we investigate how location privacy is affected by the amount of time a User is connected to a wireless local area network (WLAN). We propose a game-theoretic model to capture the interaction between a Company and a User. The former offers some services to the latter, while he is connected to a WLAN that belongs to the Company. We assume that the Company uses a wireless communication technology to localize users in order to increase its profits by launching targeted advertisements or by selling User location data to third parties. It is worth noting here that our analysis is not restricted to localization within a WLAN. It can, for instance, be rectified to increase location privacy in scenarios where phones can be tracked without using their GPS or WiFi data, e.g. by studying only their power usage over time, as in [1].

Our work is motivated by the observation that location disclosure entails different privacy risks, and we can realistically say that the location data is valuable to the Company. Suppose, for example, that the Company has established its wireless network within a shopping centre. The location data of the visitors can be utilized for:

- *optimization of stores:* the Company can optimize the store design based on heat-maps of customer movements;
- *targeted advertisements:* if the Company knows the location of customers, it can send product information based on their location, creating *location-based spam*;
- *profiling:* from the User's long-term location information, the Company can create profiles, and use them for strategic decisions, or even sell this information to third parties.

In order to obtain the desired location data, the Company establishes a passive localization system based on signal information (e.g., Received Signal Strength (RSS)) of the users' devices. During connection time, the User can be localized and therefore the more the User stays connected, the more location traces can be collected by the Company. The latter offers services to the User, which can compensate the location privacy loss. These services may include free broadband access, geolocation services, and discounts for certain products or lotteries.

This paper is organized as follows. The system model, including both the Company and the User, is described in Section II. In Section III, we formulate the Location Privacy Game (LPG) by defining the players' strategies, types, and payoffs. Section IV is dedicated to the theoretical analysis presenting the equilibria conditions of the game, and deriving the User's best response and the Company's optimal strategy in LPG. In Section V, we present the performance evaluation results, which demonstrate the effectiveness of our game-theoretic approach. The related work is discussed in Section

978-1-4673-7952-6/15 $31.00 © 2015 IEEE
DOI 10.1109/Trustcom.2015.476

VI, while Section VII concludes the paper.

II. System Model

In our model, we assume a Company which controls the communication infrastructure (CI) (e.g., WiFi network) of a building (e.g., a shopping centre) and offers services to the visitors when they are connected to CI. We consider the User as the entity that can utilize these services, and at the same time, he can be located by the Company, which leads to suffering some location privacy loss. For a list of symbols used in this paper, see Table I.

A. Passive Localization System

We assume that the Company maintains a passive indoor localization system to determine the location $\boldsymbol{l}(\tau)$ of the User at time τ. The passive localization system determines a location estimate $\boldsymbol{l}_{est}(\tau)$, which is an approximation of the User's true location at time τ. The precision of this approximation is determined by the number of data packets that the User transmits per second, i.e., the more data the User sends the more precise $\boldsymbol{l}_{est}(\tau)$ becomes; however, modeling this relationship is out of the scope of this paper. This approach is different from cases where the User actively reports his location in order to use LBSs [2], [3], as we assume that localization occurs without the User's active participation.

Any position estimate $\boldsymbol{l}_{est}(\tau)$ is biased with an error

$$l_{err}(\tau) := \left\| \boldsymbol{l}(\tau) - \boldsymbol{l}_{est}(\tau) \right\| . \tag{1}$$

As the User is moving, $l_{err}(\tau)$ can take different values (i.e., $l_{err}(\tau)$ is a random variable). We denote the expected value $\mathbb{E}[l_{err}(\tau)]$ by $\hat{l}$.

B. Location Privacy

We assume that the User is roaming within the Company's area for time $T \in \mathbb{Z}$, but his device stays connected to the CI of the Company only for time $t \in \mathbb{R} : \delta \leq t \leq T$, where δ is very small value. We have assumed here that every User needs some minimal amount of connection time δ, for example, in order to become aware of the services that the Company offers. The lower the value of t, the lower the location privacy loss of the User, as the User can be located only during t, since there are no data packets transmitted when the User is not connected. Then, the Users' location privacy, when connected to CI for time t, equals

$$p(t) := \frac{T}{t}\,\hat{l}. \tag{2}$$

In order to increase his location privacy, the Users seeks a minimum t with respect to some minimum required service level, which will define later in this section. This is based on the assumption that the longer the User stays connected, the higher level of service he receives.

C. User Types

In this paper, we assume that there are multiple User types. This is motivated by real-world scenarios where a company provides some services and several users (i.e., of different types) are roaming within its service area. The User type is determined by the User's preference to protect his location privacy. For a User of type $i \in \mathcal{A}$, where $\mathcal{A}$ is the set of User types, his privacy preferences are modeled by $\Pi_i \in [0, 1]$, which we call the *privacy factor*. For instance, $\Pi_i = 1$ models a User who completely ignores the service provided by the Company in favor of maximizing his privacy. We assume that Π_i is entered by the User on his mobile device.

TABLE I
List of Symbols

Symbol	Description
$\hat{l}$	Expected localization error
T	User visiting time
$\mathcal{A}$	Set of User types
α_i	Likelihood of the User being of type i
Π_i	Privacy factor for type i User
t_i	Connection time of type i User
p_i	Location privacy of type i User
δ	Very small value, lower bound of the connection time
S	Company's offered service level
S^*	Upper bound of Company's offered service level
$\hat{S}$	Expected service level
σ	User experienced service level
Θ	Unit service cost
Ξ	Unit service benefit
ϕ_j	Probability of the j-th service level to be chosen
$\mathcal{S}_U$	Set of User's pure strategies
$\mathcal{S}_C$	Set of Company's pure strategies
μ_i	Threshold value of the offered expected service level, where the best response strategy of type i User changes

We let t_i denote the connection time that a User of type i chooses. Thus, the location privacy of User type i, for connection time t_i, is given by $p_i = \frac{T}{t_i}\,\hat{l}$, where we denote $p_i(t_i)$ by p_i for convenience.

D. Offered and Experienced Service Level

We assume that the Company can offer a service level $S \in \mathbb{Z}$, with $0 < S \leq S^*$, to the User. The service level S represents the highest possible additive level of the offered services. We differ the User's *experienced service level* $\sigma(t, S)$ from S, and we assume that $\sigma(t, S) = S$ if and only if the User stays connected for $t = T$; otherwise, $\sigma(t, S) < S$. It is easy to see that the highest possible service level that the User can experience equals S^*, and it can be obtained only when the Company offers S^* and the User chooses $t = T$.

The experienced service level $\sigma(t, S)$ is modeled as a linear non-decreasing function. In practice, $\sigma(t, S)$ is discrete (i.e., the Company gives out a discount or not). Therefore, $\sigma(t, S)$ gets a connection time t and an offered service level, and it provides an attainable discrete service level as follows

$$\sigma(t, S) := \frac{t}{T}\,S. \tag{3}$$

III. Location Privacy Game

In this section, we define the Location Privacy Game (LPG), which is a 2-player Bayesian Stackelberg game between the Company C and the User U. In the LPG, the leader (Company) first commits to his strategy, which is observed by the follower (User). The Bayesian extension to the Stackelberg game allows us to capture multiple types of followers, where each follower has its own payoff values. We denote the set of User types by $\mathcal{A}$, and the User is of type i with probability α_i, decided by Nature [4].

A. Strategies

In the LPG, the Company decides upon the offered service level S with knowledge of the probability distribution over the different User types. On the other hand, the User wants to consume some of these services while respecting his location privacy preferences. The Company advertises S, and the User can observe this and play his best response by choosing an optimal t. The Company wishes that the User will stay connected for as long as possible, and therefore, to be able to construct the entire path that the User has followed; however, each offered service level has a cost, which increases with S. This cost is modeled by the monotonically increasing function $\Theta\, S$, where Θ is a positive constant called the *unit service cost*. We also assume that the Company benefits from tracking the User's location, for example, by selling his location data to third parties. We model the Company's benefit as a monotonically increasing function of t, which is given by $\Xi\, \frac{1}{p(t)}$, where Ξ is a positive constant called the *unit service benefit*.

The pure strategy choice of the Company is to offer a service level S, and we express its strategy set as $\mathcal{S}_C := \{1, \ldots, S^*\}$. We also express the set of the User's pure strategies as $\mathcal{S}_U := [\delta, T]$. Note that, for the remainder of this paper, we will denote the j-th service level by S_j, and the connection time chosen by a User of type i is denoted by t_i, as mentioned earlier.

A player's mixed strategy is a distribution over the set of his pure strategies. For the Company, the canonical representation of its mixed-strategy space is a discrete probability distribution over the set $\mathcal{S}_C$. We represent a mixed strategy of the Company as an $|\mathcal{S}_C|$-dimensional vector $\mathbf{\Phi}$, where ϕ_j is the probability of offering the j-th service level. In the LPG, we assume that the User plays only pure strategies, since there always exists a pure strategy that is a best response for the User, as it is also explained in [5].

B. Payoffs

1) Company: For a given User type i and strategy profile $(\mathbf{\Phi}, t_i)$, the Company's payoff is

$$\begin{aligned} \mathcal{U}_C^{(i)}(\mathbf{\Phi}, t_i) &:= \Xi\, \frac{1}{p_i} - \Theta \sum_{j \in \mathcal{S}_C} \phi_j\, S_j \\ &= \frac{\Xi}{T}\, \frac{1}{\hat{l}}\, t_i - \Theta \sum_{j \in \mathcal{S}_C} \phi_j\, S_j. \end{aligned} \tag{4}$$

This payoff is in the form $\Psi\, t_i - \Theta \sum_{j \in \mathcal{S}_C} \phi_j\, S_j$, where $\Psi, \Theta,$ are positive constants, and

$$\Psi = \frac{\Xi}{T}\, \frac{1}{\hat{l}}. \tag{5}$$

The *overall expected payoff* of the Company is a weighted combination of its expected payoff against all user types. We represent the Users' strategies, one per each type, as an $|\mathcal{A}|$-dimensional vector $\mathbf{t} = [t_i]$, where $t_i \in \mathcal{S}_U$. Then, from Eq. (4), we have that the Company's overall expected payoff is

$$\begin{aligned} \mathcal{U}_C(\mathbf{\Phi}, \mathbf{t}) &= \sum_{i \in \mathcal{A}} \alpha_i \cdot \mathcal{U}_C^{(i)}(\mathbf{\Phi}, t_i) \\ &= \sum_{i \in \mathcal{A}} \alpha_i \left[\Psi\, t_i - \Theta \sum_{j \in \mathcal{S}_C} \phi_j\, S_j \right]. \end{aligned} \tag{6}$$

2) User: For a given offered service level S and connection time t_i, the User's payoff is determined by both the achieved privacy and the experienced service level as follows:

$$\begin{aligned} \mathcal{U}_U^{(i)}(S, t_i) &:= \Pi_i\, p_i + (1 - \Pi_i)\, \sigma(t_i, S) \\ &= \Pi_i\, T\, \hat{l}\, \frac{1}{t_i} + (1 - \Pi_i)\, \frac{1}{T}\, S\, t_i, \end{aligned} \tag{7}$$

which is in the form $\Psi_1\, \frac{1}{t_i} + \Psi_2\, S\, t_i$, where Ψ_1, Ψ_2 are positive constants, for a specific User type i, and

$$\begin{cases} \Psi_1 = \Pi_i\, T\, \hat{l} \\ \Psi_2 = (1 - \Pi_i)\, \frac{1}{T}. \end{cases} \tag{8}$$

Hence, the User's payoff for a mixed strategy $\mathbf{\Phi}$ of the Company is

$$\mathcal{U}_U^{(i)}(\mathbf{\Phi}, t_i) = \Psi_1\, \frac{1}{t_i} + \Psi_2\, t_i \sum_{j \in \mathcal{S}_C} \phi_j\, S_j. \tag{9}$$

It is easy to see that there is a trade-off between location privacy and experienced service quality level when choosing t. For instance, staying connected for long time leads to high σ but low p, and vice versa.

IV. Analysis

In the analysis, our goal will be to find the User's best response and the Company's optimal strategies, which are defined as follows.

Definition 1: A User strategy is a *best response* if it maximizes the User's payoff, taking the Company's offered service level as given.

The standard solution concept for Stackelberg games is the *Strong Stackelberg Equilibrium* (SSE) [6].

Definition 2: At the Strong Stackelberg Equilibrium (SSE) of the LPG

1) for every type i, the User of type i plays a best-response t^* to any Company strategy $\mathbf{\Phi}$, that is,

$$\mathcal{U}_U^{(i)}(\mathbf{\Phi}, t^*) \geq \mathcal{U}_U^{(i)}(\mathbf{\Phi}, t),\ \forall\, t \neq t^*;$$

2) the Users break ties in favor of the Company, that is, when there are multiple best responses to a Company strategy $\mathbf{\Phi}$, the Users play the best responses $\mathbf{t}^*$ that maximize the Company's payoff:

$$\mathcal{U}_C(\mathbf{\Phi}, \mathbf{t}^*) \geq \mathcal{U}_C(\mathbf{\Phi}, \mathbf{t}),\ \forall\, \mathbf{t} \text{ best response};$$

3) the Company plays a best-response $\mathbf{\Phi}^*$, which maximizes its payoff given that the Users' strategies are given by the first two conditions (i.e., Users always play best responses with tie-breaking in favor of the Company):

$$\mathcal{U}_C(\mathbf{\Phi}^*, \mathbf{t}^*(\mathbf{\Phi})) \geq \mathcal{U}_C(\mathbf{\Phi}, \mathbf{t}^*(\mathbf{\Phi})), \, \forall \, \mathbf{\Phi},$$

where $\mathbf{t}^*(\mathbf{\Phi})$ denotes the Users best responses with tie-breaking to a Company strategy $\mathbf{\Phi}$.

Note that, in our game, the tie-breaking rule has no practical implications, it merely eliminates some pathological mathematical cases where the game would have no equilibrium otherwise.

Since the Company's equilibrium strategies maximize its payoff, given that Users maximize their own payoffs, we will refer to them as optimal strategies for the remainder of the paper.

Definition 3: A Company strategy is *optimal* if it maximizes the Company's payoff given that the User will always play a best-response strategy with tie-breaking in favor of the Company.

A. Representing the Company's Mixed Strategies

First, observe that both the Company's and the User's expected payoffs depend on the Company's mixed strategy $\mathbf{\Phi}$ only through the expected service level $\sum_{j \in \mathcal{S}_C} \phi_j \, S_j$. To simplify our analysis, we now introduce $\hat{S}$ to denote the expected service level. For any mixed strategy $\mathbf{\Phi}$ of the Company, we can compute the corresponding $\hat{S}$ as $\hat{S} = \sum_{j \in \mathcal{S}_C} \phi_j \, S_j$. Then, we can express the Company's expected payoff as

$$\mathcal{U}_C(\hat{S}, \mathbf{t}) = \sum_{i \in \mathcal{A}} \alpha_i \Big[\Psi \, t_i - \Theta \, \hat{S} \Big] \tag{10}$$

and the User's expected payoff as

$$\mathcal{U}_U^{(i)}(\hat{S}, t_i) = \Psi_1 \, \frac{1}{t_i} + \Psi_2 \, t_i \, \hat{S}. \tag{11}$$

Furthermore, it is also clear that, for any $\hat{S} \in [\min_{j \in \mathcal{S}_C} S_j, \max_{j \in \mathcal{S}_C} S_j]$, there exists a mixed strategy $\mathbf{\Phi}$ for the Company such that $\sum_{j \in \mathcal{S}_C} \phi_j \, S_j = \hat{S}$. Hence, we can use $\hat{S} \in [\min_{j \in \mathcal{S}_C} S_j, \max_{j \in \mathcal{S}_C} S_j]$ to represent the Company's mixed strategies, and the problem of finding an optimal strategy reduces to finding an optimal $\hat{S}$ value.

B. User's Best Response

In order to find an optimal strategy for the Company, we first have to characterize the Users' best-response strategies.

Lemma 1: For any Company strategy $\hat{S}$, the User's best response is either δ or T.

Proof: The domain of the payoff function $\mathcal{U}_U^{(i)}(\hat{S}, t_i)$ is t_i in $[\delta, T]$. Then, we can compute the first derivative of $\mathcal{U}_U^{(i)}(\hat{S}, t_i)$ with respect to t_i as $\frac{\partial \mathcal{U}_U^{(i)}}{\partial t_i} = -\Psi_1 \frac{1}{t_i^2} + \Psi_2 \, \hat{S}$. Next, we can compute the second derivative of $\mathcal{U}_U^{(i)}(\hat{S}, t_i)$ with respect to t_i as $\frac{\partial^2 \mathcal{U}_U^{(i)}}{\partial t_i^2} = 2 \, \Psi_1 \frac{1}{t_i^3} + 0 > 0$. Since the second derivative is always positive on $[\delta, t_i]$, we have that the payoff function $\mathcal{U}_U^{(i)}(\hat{S}, t_i)$ is a convex function of t_i. It follows from the convexity of the function that the maximum payoff is attained at one of the endpoints δ and T. Therefore, the User's best response is either δ or T. ■

Theorem 1: If User of type i plays a best-response strategy and breaks ties in favor of the Company, then his strategic choice for a Company strategy $\hat{S}$ is

- $t_i = \delta$ if $\hat{S} < \mu_i$,
- $t_i = T$ if $\hat{S} \geq \mu_i$,

where

$$\mu_i = \frac{\Psi_1}{\Psi_2} \, \frac{1}{\delta \, T}. \tag{12}$$

The above theorem basically shows that the User's best response is a non-decreasing right-continuous step function of $\hat{S}$ (see Fig. 1 for an illustration). Note that, if the threshold μ_i is outside the interval $[\min_{j \in \mathcal{S}_C} S_j, \max_{j \in \mathcal{S}_C} S_j]$, then the best response is constant.

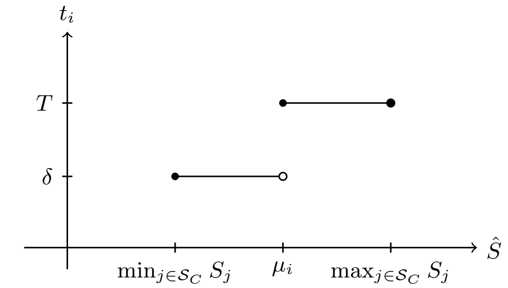

Fig. 1. Illustration of the User's best response with tie-breaking as a function of the Company's strategy $\hat{S}$.

Proof: From Lemma 1, we have that the User's strategic choice is either δ or T. Since the Company's payoff is always an increasing function of t_i, the User has to choose T if both δ and T are best responses, as the User breaks ties in favor of the Company. Hence, it remains to characterize the case when δ is the only best response. The strategy δ is a better response than the strategy T if and only if

$$\begin{aligned} &\mathcal{U}_U^{(i)}(\hat{S}, \delta) > \mathcal{U}_U^{(i)}(\hat{S}, T) \Rightarrow \frac{\Psi_1}{\delta} + \Psi_2 \, \delta \, \hat{S} > \frac{\Psi_1}{T} + \Psi_2 \, T \, \hat{S} \\ &\Rightarrow \Psi_1 \, (\frac{1}{\delta} - \frac{1}{T}) > \Psi_2 \, \hat{S} \, (T - \delta) \Rightarrow \Psi_1 \, \frac{T - \delta}{\delta \, T} > \\ &\Psi_2 \, \hat{S} \, (T - \delta) \Rightarrow \hat{S} < \frac{\Psi_1}{\Psi_2} \, \frac{1}{\delta \, T}. \end{aligned} \tag{13}$$

■

C. Company's Optimal Strategy

Lemma 2: Suppose that we are given a set of User strategies $\mathbf{t} = (t_1, t_2, \ldots, t_{\mathcal{S}_U})$, and the Company's strategy space is limited to $\hat{S}$ values for which $\mathbf{t}$ is a best response. Then, the Company's payoff is a strictly decreasing function of $\hat{S}$.

Proof: We can reformulate the Company's payoff function as

$$\begin{aligned} \mathcal{U}_C(\hat{S}) \quad &= \textstyle\sum_{i \in \mathcal{A}} \alpha_i \Big[\Psi \, t_i - \Theta \, \hat{S} \Big] \\ &= \underbrace{-\Theta}_{<0} \, \hat{S} + \underbrace{\sum_{i \in \mathcal{A}} \alpha_i \Big[\Psi \, t_i \Big]}_{\text{constant}}. \end{aligned} \tag{14}$$

Hence, on this limited strategy space, the Company's payoff is a strictly decreasing function of $\hat{S}$. ■

Theorem 2: The Company's optimal strategy is either $\min_{j \in \mathcal{S}_C} S_j$ or one of the threshold values μ_i defined in Theorem 1.

Proof: The Users' threshold values $\mu_1, \mu_2, \ldots, \mu_{|\mathcal{S}_U|}$ divide the Company's strategy space $[\min_{j \in \mathcal{S}_C} S_j, \max_{j \in \mathcal{S}_C} S_j]$ into at most $|\mathcal{S}_U| + 1$ contiguous intervals. From Lemma 2, we have that the Company's payoff is strictly decreasing on each one of these intervals. From Lemma 1, we have that each of these intervals is left-closed (see Fig. 2 for an illustration). Therefore, the Company's payoff attains its maximum at one of the left endpoints, that is, either at $\min_{j \in \mathcal{S}_C} S_j$ or at one of the threshold values μ_i. ■

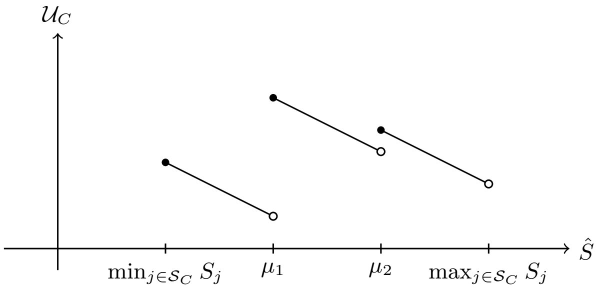
$\mathcal{U}_C$
$\hat{S}$
$\min_{j \in \mathcal{S}_C} S_j$
μ_1
μ_2
$\max_{j \in \mathcal{S}_C} S_j$

Fig. 2. Illustration of the Company's expected payoff as a function of its strategy $\hat{S}$. In this figure, the optimal strategy is μ_1.

V. Results

For the purposes of this section, we have used a wireless (IEEE 802.11) localization testbed to derive realistic expected localization error $\hat{l}$ values, which we have then used to derive the payoffs of the Company and the User. We have undertaken simulations to compare the payoffs of different User types. Additionally, we have compared the Bayesian Company strategy with a strategy that assumes that all the Users have the same average Π_i value. Finally, we have demonstrated the benefit of our game-theoretic solution as opposed to non-strategic decisions.

For this case study, we define the set of possible expected service levels as $\{1, 2, \ldots, 10\}$. Since LPG is a Stackelberg game, the User is aware of these service levels and he chooses the one that maximizes his payoff. On the other hand, the Company chooses an optimal $\hat{S} \in \{1, 2, \ldots, 10\}$. In our testbed, the measurement stations (MSs) are devices that use the IEEE 802.11 protocol (i.e., WiFi) and their wireless cards are set into monitor mode. We performed practical measurements by using an IEEE 802.11 testbed. We have generated Received Signal Strength (RSS) values as inputs to our localization algorithm. To generate these values, we use the formula [7]

$$P_{R_i} = P_0(d_0) - 10\, n_i \, \log_{10} \frac{d_i}{d_0} + X, \text{ where} \tag{15}$$

- P_{R_i} is the received power at station i;
- $P_0(d_0)$ is a reference power measured at distance d_0;
- n_i is the path loss exponent, which depends on the environment between User and measurement station i;
- d_i is the distance between MS i and User's device;
- X is a zero-mean log-normal distributed random variable reflecting the flat fading with standard deviation ϵ_X.

We have used a Nexus 4 mobile device, which sends 1000 packets per second, and we have selected twelve locations where the User could be. We have taken 1000 measurements at each of these locations, for 4 directions, resulting in 4000 measurements for each location. By averaging these measurements we have derived $n_i = 0.75 \ \forall i$, $d_0 = 0.7$ meters, $P_0(d_0) = -59$, and $\epsilon_X = 1$. We use the previously identified values and (15) to simulate and derive a mean localization error when different number of packets are sent by the User device. The latter affect the localization error because of the flat fading X. Therefore, we use 1000 random locations from the interval $[0, 10] \times [0, 10]$ and locate the User using multilateration [8, p. 164]. We assume three different values 1000, 500, and 200 for the amount of data sent by a device resulting in the mean localization errors 40.12m, 46.64m, and 58.04m, correspondingly. The errors depend strongly on the environment and obstacles (e.g., moving people, walls) in the propagation path. We also see that multilateration does not perform well at all. However, the performance of this localization system falls out of the scope of this paper.

Following the results of Westin [9], we classify the users into the following three categories: Privacy Fundamentalists (PFs); Privacy Unconcerned (PUs); and Privacy Pragmatists (PPs). According to [9], PFs "*reject the consumer-benefit or societal-protection claims for data uses and sought legal-regulatory privacy measure*;" PUs are "*ready to supply their personal information to business and government and reject what is seen as too much privacy fuss*;" and PPs "*examine the benefits to them of the data collection and use, want to know the privacy risks and how organizations propose to control those, and then decide whether to trust the organization or seek legal oversight*." Therefore, we define the set of User types as $\mathcal{A} = \{\text{PU}, \text{PP}, \text{PF}\}$, and we set their corresponding privacy factors Π_i as $\{0.2, 0.5, 0.8\}$. We have simulated a scenario where the User's minimal connection time is $\delta = 2$, and the unit service benefit Ξ is 50% higher than the unit service cost Θ. Note that the above privacy factor values have been chosen for the purpose of evaluating our model and they should not be considered as a recommendation from the literature. We also recognize that in real-life scenarios we might notice the "privacy paradox", according to which people tend to express extreme privacy preferences but act differently, in a rather erratic way. However, in our work here, we assume that users are rational entities whose actions are consistent with their privacy preferences.

Fig. 3 shows the Company's payoff for the different mean localization error values, as discussed previously. We notice that for $\hat{l} = 40.12$m, $\mathcal{U}_C$ becomes negative when $T = 17$, and for both $\hat{l} = 46.64$m and $\hat{l} = 58.04$m, when $T = 7$. These low values of total connection time demonstrate the need for an effective localization system, if the Company decides to

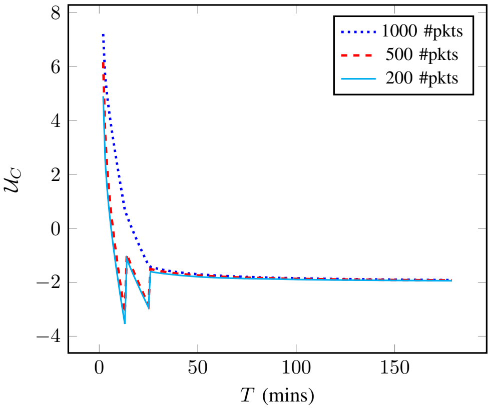

Fig. 3. Company's payoff for different number of packets sent by the User device.

implement the model discussed in this paper.

For the remainder of this paper, we assume that $\hat{l} = 2$m. In Fig. 4 we compare the payoffs arising from the optimal Bayesian Company strategy and from the optimal "Averaging strategy." Both strategies are evaluated in the Bayesian model, assuming that User types are uniformly distributed. The former strategy considers the differences between the User types, and as a result, correctly assumes that the privacy factor Π_i is drawn from $\{0.2, 0.5, 0.8\}$ uniformly at random. On the other hand, the Averaging strategy assumes that the users are homogeneous and that the privacy factor always takes its expected value 0.5 (i.e., assumes a single User type which has the average Π value 0.5). This comparison allows us to determine how much the Company can gain from knowing the actual distribution of the User types. For visiting time $T = 84$ minutes, the Company's payoff decreases with T for both strategies. However, we notice that the Bayesian Company strategy outperforms the Averaging strategy when $T > 36$. Furthermore, the Averaging and Bayesian strategies give negative payoffs for $T > 51$, and $T > 82$ correspondingly. Given that when negative payoffs are reached the Company must rather decide not to provide any services, the Bayesian strategy gives 31 minutes extra time for the Company to make profit.

More importantly, in Fig. 5 we show how the Company benefits from following the Stackelberg strategy as opposed to non-strategic decisions, such as the maximum $\hat{S}$ value 10, the minimum $\hat{S}$ value 1, and also the weighted $\hat{S}$ value. The latter is given by first assuming that the Company chooses as expected service levels $[2, 5, 8]$ when the User privacy factors are $[0.2, 0.5, 0.8]$. Secondly, the Company multiplies each $\hat{S}$ value by the probability α_i of a User being of type i.

Following the results of [9], we have used the probability distribution $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \alpha_3] = [0.2, 0.55, 0.25]$ over $\mathcal{A}$ and, therefore, over $\{0.2, 0.5, 0.8\}$ for the Bayesian model. We assume that the Company is aware of $\boldsymbol{\alpha}$. Thus, the Company can compute its optimal expected service level, by using Eq. (10). It is easy that, for the weighted strategy, given $\boldsymbol{\alpha}$, we have that $\hat{S} = 0.2 \cdot 2 + 0.55 \cdot 5 + 0.25 \cdot 8 = \lfloor 5.15 \rfloor = 5$.

First, we notice that for all of the $\hat{S}$ values, the Company's payoff is a decreasing function of the User's visiting time T in this Bayesian model. More specifically, the results show that if the Company chooses the Max strategy, its payoff becomes negative for $T > 10$, while for the Min strategy, the Company can keep providing services for 21 extra minutes ($T = 31$), before its payoff becomes negative. This time is improved by 20 minutes when the Company chooses the weighted value, leading to $T = 51$ before its payoff becomes negative. The best performance is achieved when the Company chooses the $\hat{S}$ determined by the Strong Stackelberg Equilibrium of the LPG. This allows the Company to make profit (i.e., having positive payoff), for 57 minutes. Although the 6 extra minutes improvement of the Company's payoff per User is not remarkable, we must note that such an improvement leads to significantly higher Company profits when considering a high number of users, as in realistic scenarios.

It is also worth noting that for $T > 51$, the Company's payoff decreases significantly (reaching -6.529), when the weighted value is chosen. On the other hand, although the Company's payoff becomes negative for $T > 57$, its value remains -0.0184 for the rest of the simulated time. This can be useful if we assume that the Company occasionally decides not to stop offering services immediately after its payoff becomes negative, in favor of its Users.

Besides investigating the Company's payoff, we have looked into the payoffs of different User types, when the User plays his best response according to Definition 2. In Fig. 6, we have plotted these payoffs for the same parameters Θ, Ξ, $\boldsymbol{\alpha}$ as in the above results, and different visiting times T. We observe that for a Privacy Fundamentalist (PF), the payoff increases as a function of the visiting time, even from the very first minute. In contrast, the payoff of a Privacy Pragmatist (PP) User equals 0 for visiting times less than 60 minutes. For higher values than this, the User payoff becomes positive taking the value 3.0. Thereafter, for $T > 60$ PP's payoff only increases. Finally, the payoff of a Privacy Unconcerned (PU) User, remains 0 for the visiting time values lower than 74. At this point, the payoff

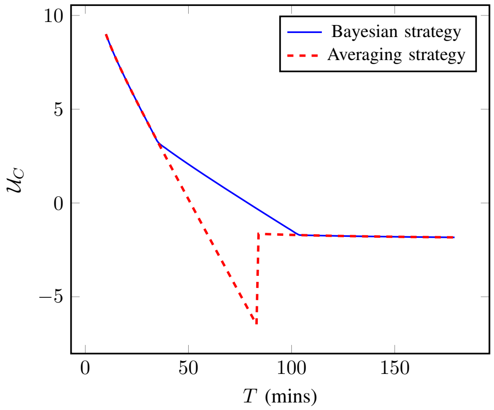

Fig. 4. Comparing a Bayesian with an Averaging strategy for the Company.

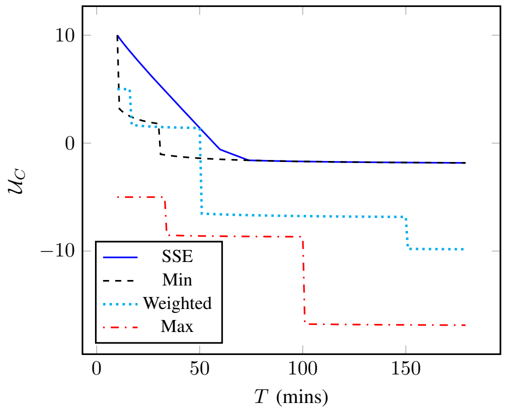

Fig. 5. Comparing the payoff of the Company for different non-strategic decisions and the strategy at the Strong Stackelberg Equilibrium (SSE).

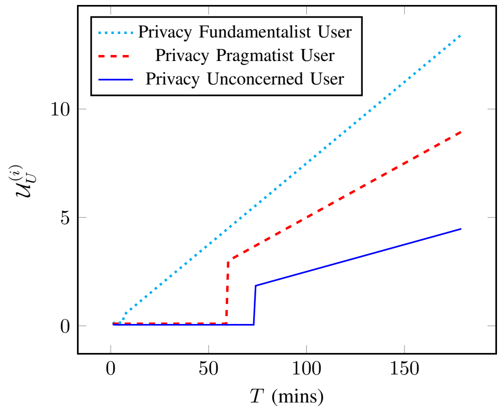

Fig. 6. Payoffs of the different User types at the SSE of the LPG.

becomes 1.85, and thereafter it only increases. However, it remains lower than the PP's payoff for the rest of the time. Note that, both PP's and PU's payoffs are lower than the PF's payoff at all times, highlighting that the latter is the most favored User type in our model.

Finally, in Fig. 7, we can see the thresholds for the different User types as a function of the visiting time. As expected, the results show that for all User types, threshold values increase with the visiting time. This means that the higher the visiting time T, the higher the expected service level $\hat{S}$ must be for the User to stay connected for T, as opposed to remaining connected for a small δ. This happens because the User is more concerned about his location privacy for longer visiting periods; therefore, he has to receive a higher $\hat{S}$ in order to consider it worthwhile (i.e., best response) to be connected to the CI of the Company for T.

Given that $\hat{S} \in \{1, 2, \ldots, 10\}$, the results show that a PF User will not get connected for more than δ minutes when the visiting time T exceeds 34 minutes, regardless of the expected service level $\hat{S}$, offered by the Company. Likewise, a PP User considers staying connected to the Company's CI for the whole visiting time, for T values only up to 101 minutes, if the required $\hat{S}$ is offered. Lastly, a PU User can stay connected for the maximum simulation time $T = 180$, for the "right" $\hat{S}$ value. To have a more clear view on how quickly $\hat{S}$ must increase to satisfy the requirements of the different User types, we have derived the slope of each threshold function for each User type. For this derivation, we have computed the derivative of μ_i with respect to T. Thus, from (12), we have $\frac{\partial \mu_i}{\partial T} = \frac{\partial \frac{\Psi_1}{\Psi_2} \frac{1}{\delta T}}{\partial T} = \frac{\Psi_1}{\Psi_2} \frac{1}{\delta} = \frac{\Pi_i \, \hat{l}}{(1-\Pi_i)\, \delta}$. From this, we found the following slope values: PU: 0.033, PP: 0.1, and PF: 0.3. We notice that, for the same visiting time, a PF User requires a 3 times higher $\hat{S}$ offered than a PP User, in order to stay connected for T, and 9 times higher $\hat{S}$ offered than a PU User.

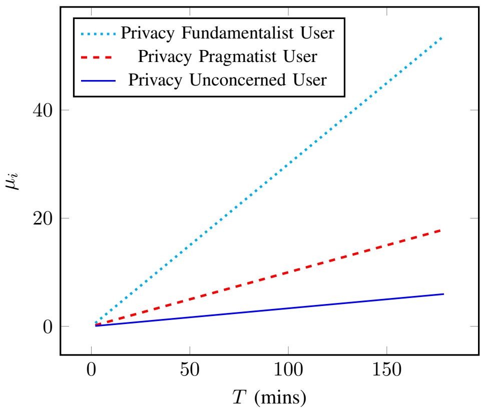

Fig. 7. Threshold values, in terms of offered expected service levels, where the User's best response changes, for the different User types.

VI. Related Work

In this section we discuss state-of-the-art work at the intersection of game theory and location privacy. A thorough survey related to this has been published by Manshaei et al. [2]. The majority of the papers model two players: the attacker and the user.

According to [10], in order to design an optimal privacy-protection mechanism it is crucial to take the *knowledge of an attacker* into consideration. This means, for example, that the adversary is aware of the utilized location protection algorithm and the access profile of the user i.e., the probability distribution describing the user access to Location-based services (LBS) in a certain region. This assumes, that the user contacts the LBS sporadically.

Shokri et al. [10] provide a framework to methodologically integrate this knowledge by using a zero-sum Bayesian Stackelberg game in order to derive the optimal protection strategy. In their scenario, the user is the leader and the adversary is the follower. They build on the *correctness* metric explained above to measure the users' location privacy. Their game consists of four steps. First, the Nature selects a location r for the user to access the LBS. Second, the user protects his/her location by creating a pseudo-location r' with a function f. Third, the attacker observes r' and tries to infer r using the knowledge of f and the access profile of the user resulting in an estimation

$\hat{r}$. Finally, the adversary pays an amount $d(r, \hat{r})$ to the user. Here $d(\cdot)$ is a distance function and represents the estimation error of the adversary. The authors derive optimal strategies for both, user and adversary.

Furthermore, Shokri et al. [11] present a privacy preserving approach relying on user-collaboration. Their solution, called *MobiCrowd*, requires the mobile devices to communicate wirelessly and in a peer-to-peer manner. The mobile devices keep their context information in a buffer, until it expires, and they pass it to other collaborative users seeking such information. This leads to less communication with the service provider because a user contacts the provider only if there are no other users, with the requested information, in range. In this initial work no game theory is used but it is the basis for [12], where Santos et al. extend their work by analyzing the collaborative behavior of users in MobiCrowd with game-theoretic methods. The two Nash game equilibria, which they have derived, favor mutual cooperation and mutual defection. In a second game they combine game theoretic analysis with an epidemic model to investigate the behavior of more than two users. In this way, they derive the optimal threshold α_{opt} for cooperation that optimizes the payoff of a user.

Chorppath and Alpcan [13] establish a privacy mechanism-design game between a company and its mobile users. The company offers incentives to the users in order them to report their location with a certain level of accuracy. The authors derive the total budget that a company must invest on providing incentives to obtain a desired minimum level of location accuracy from all the users.

As far as we know, all above papers make the assumption that users actively report their location in order to use LBSs. They also look into anonymity issues, and they aim to decouple the user identity from his location. However, modern devices do not come with the capability of changing, for instance, their users MAC address, and therefore confusing the attacker about their real identity.

To the best of our knowledge, our work is the first game-theoretic approach investigating users' strategies in a passive localization environment, where location is derived by raw signal measurements, and the only parameter that the user can control is the amount of connection time. Finally, our work is innovative, as it is the first one to model the economically-rational decision-making of the service provider in conjunction with rational decision-making of the users who wish to protect their location privacy.

VII. Conclusion

This paper presents a game-theoretic model, in which a Company incentivizes a User to permit location tracking, by offering "attractive" service levels based on the different User types. The User's location is tracked by a passive localization system, which is established and maintained by the Company. We have defined a Stackelberg game, called the Location Privacy Game (LPG), according to which a User selects the amount of time he stays connected (i.e., connection time) to the Company's network (e.g., WiFi), and the Company chooses the level of services that are offered to the User. We have presented theoretical results characterizing the equilibria of the game. Then, we have developed an IEEE 802.11 wireless testbed, which facilitated the computation of different expected localization errors for a User who is equipped with a mobile device. We have used these values in our simulations to demonstrate the superiority of the game-theoretic strategy as opposed to non-strategic methods. More importantly, we have considered different User privacy types, as published by Westin [9], and have determined the service level that must be provided by the Company to incentivize the User to stay connected as long as possible to the Company's network.

Regarding our plans for future work, an interesting and actively explored research direction is developing information theory-based metrics for quantifying user privacy. Therefore modeling user and service provider decision processes using privacy games and by integration of such metrics is another direction that immediately follows. Furthermore, plans include a model extension that will facilitate user privacy within the realm of the Internet-of-Things, where localization capabilities are more often the case than the exception.

References

[1] Y. Michalevsky, G. Nakibly, A. Schulman, and D. Boneh, "Powerspy: Location tracking using mobile device power analysis," in *arXiv*, 2015.

[2] M. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J. Hubaux, "Game theory meets network security and privacy," *ACM Comput. Surv.*, no. 45, pp. 25:1–25:39, 2013.

[3] R. Shokri, G. Theodorakopoulos, P. Papadimitratos, E. Kazemi, and J. Hubaux, "Hiding in the mobile crowd: Location privacy through collaboration," *IEEE Transactions on Dependable and Secure Computing*, vol. 11, no. 3, pp. 266–279, 2013.

[4] T. Alpcan and T. Basar, *Network security: A decision and game-theoretic approach*. Cambridge University Press, 2010.

[5] M. Jain, C. Kiekintveld, and M. Tambe, "Quality-bounded solutions for finite Bayesian Stackelberg games: Scaling up," in *Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, vol. 3, 2011, pp. 997–1004.

[6] Z. Yin, D. Korzhyk, C. Kiekintveld, V. Conitzer, and M. Tambe, "Stackelberg vs. Nash in security games: Interchangeability, equivalence, and uniqueness," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, vol. 1, 2010, pp. 1139–1146.

[7] A. Bahillo, S. Mazuelas, R. M. Lorenzo, P. Fernández, J. Prieto, R. J. Durán, and E. J. Abril, "Hybrid RSS-RTT localization scheme for indoor wireless networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, 2010.

[8] A. Bensky, *Wireless Positioning Technologies and Applications*, ser. GNSS Technology and Applications Series. Artech House, 2008.

[9] A. Westin, "Social and political dimensions of privacy," *Journal of Social Issues*, vol. 59, no. 2, pp. 431–453, 2003.

[10] R. Shokri, G. Theodorakopoulos, C. Troncoso, J. Hubaux, and J. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proceedings of the 2012 ACM Conference on Computer and communications Security (CCS)*. ACM, 2012, pp. 617–627.

[11] R. Shokri, P. Papadimitratos, G. Theodorakopoulos, and J. Hubaux, "Collaborative location privacy," in *Proceedings of the 2011 IEEE 8th International Conference on Mobile Adhoc and Sensor Systems (MASS)*. IEEE, 2011, pp. 500–509.

[12] F. Santos, M. Humbert, R. Shokri, and J. Hubaux, "Collaborative location privacy with rational users," in *Proceedings of the 2nd International Conference on Decision and Game Theory for Security (GameSec)*. Springer, 2011, pp. 163–181.

[13] A. K. Chorppath and T. Alpcan, "Trading privacy with incentives in mobile commerce: A game theoretic approach," *Pervasive and Mobile Computing*, vol. 9, no. 4, pp. 598 – 612, 2013.