Quantifying the Effect of Co-location Information on Location Privacy

Alexandra-Mihaela Olteanu[1], Kévin Huguenin[1], Reza Shokri[2], and Jean-Pierre Hubaux[1]

[1] School of Computer and Communication Sciences, EPFL, Switzerland
{alexandramihaela.olteanu,kevin.huguenin,jean-pierre.hubaux}@epfl.ch
[2] Department of Computer Science, ETH Zurich, Switzerland
reza.shokri@inf.ethz.ch

Abstract. Mobile users increasingly report their *co-locations* with other users, in addition to revealing their locations to online services. For instance, they tag the names of the friends they are with, in the messages and in the pictures they post on social networking websites. Combined with (possibly obfuscated) location information, such co-locations can be used to improve the inference of the users' locations, thus further threatening their location privacy: as co-location information is taken into account, not only a user's reported locations and mobility patterns can be used to localize her, but also those of her friends (and the friends of their friends and so on). In this paper, we study this problem by quantifying the effect of co-location information on location privacy, with respect to an adversary such as a social network operator that has access to such information. We formalize the problem and derive an optimal inference algorithm that incorporates such co-location information, yet at the cost of high complexity. We propose two polynomial-time approximate inference algorithms and we extensively evaluate their performance on a real dataset. Our experimental results show that, even in the case where the adversary considers co-locations with only a single friend of the targeted user, the location privacy of the user is decreased by up to 75% in a typical setting. Even in the case where a user does not disclose any location information, her privacy can decrease by up to 16% due to the information reported by other users.

Keywords: Location privacy, co-location, statistical inference, social networks.

1 Introduction

Increasingly popular GPS-equipped mobile devices with Internet connectivity allow users to enjoy a wide range of online location-based services while on the go. For instance, mobile users can search for nearby points of interest and get directions, possibly in real time, to their destinations. Location-based services raise serious privacy concerns as a large amount of personal information can be inferred from a user's whereabouts. The research community has extensively

E. De Cristofaro and S.J. Murdoch (Eds.): PETS 2014, LNCS 8555, pp. 184–203, 2014.

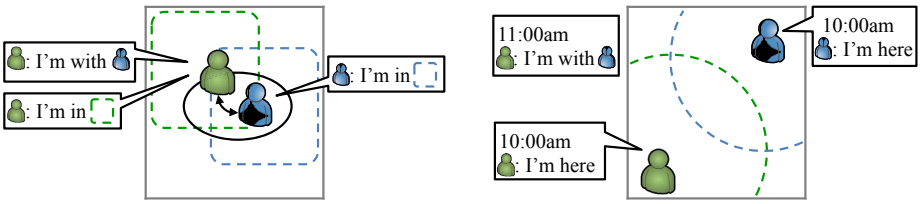


Fig. 1. Examples showing how co-location information can be detrimental to privacy. (a) A user reports being in a given area, and a second user reports being in another (overlapping) area and that she is co-located with the first user. By combining these pieces of information, an adversary can deduce that both users are located in the intersection of the two areas, thus narrowing down the set of possible locations for both of them. (b) Two users (initially apart from each other, at 10am) declare their exact individual location. Later (at 11am), they meet and report their co-location without mentioning where they are. By combining these pieces of information, the adversary can infer that they are at a place that is reachable from both of the initially reported locations in the amount of time elapsed between the two reports.

studied the problem of location privacy; more specifically, location-privacy protection mechanisms (so-called LPPMs), that can anonymize and obfuscate the users' locations before sending them to online location-based services, have been proposed [16]. In addition, formal frameworks to quantify location privacy in the case where users disclose their (possibly obfuscated) locations have been proposed [19, 20]. In such frameworks, the mobility profiles of the users play an important role in the inference of the users' locations, namely in a localization attack.

In parallel, social networks have become immensely popular. Every day, millions of users post information, including their locations, about themselves, but also about their friends. An emerging trend, which is the focus of this paper, is to report co-locations with other users on social networks, *e.g.,* by tagging friends on pictures they upload or in the messages they post. Our preliminary survey involving 132 Foursquare users, recruited through Amazon Mechanical Turk, reveals that 55.3% of the participants do report co-locations in their check-ins and that for the users who do so, on average, 2.84%±0.06 of their check-ins do contain co-location information. In fact, co-location information can be obtained in many different ways, such as automatic face recognition on pictures (which can contain the time and location at which the picture was taken in their EXIF data), Bluetooth-enabled device sniffing and reporting neighboring devices. Similarly, users who connect from the same IP address are likely to be attached to the same Internet access point, thus providing evidence of their co-location.

Attacks exploiting both location and co-location information (as mentioned in [22]) can be quite powerful, as we show in this paper. Figure 1 depicts and describes two example situations in which co-location can improve the performance of a localization attack, thus degrading the location-privacy of the users

involved. At the same time, it is clear that the proper exploitation of such information by an attacker can be complex because he has to consider jointly the (co-)location information collected about a potentially large number of users.

This family of attacks and their complexity is precisely the focus of this paper. More specifically, we make the following three contributions. (1) We identify and formalize the localization problem with co-location information, we propose an optimal inference algorithm and analyze its complexity. We show that, in practice, the optimal inference algorithm is intractable due to the explosion of the state space size. (2) We describe how an attacker can drastically reduce the computational complexity of the attack by means of well-chosen approximations. We present two polynomial-time heuristics, the first being based on a limited set of considered users and the second relying on an independence approximation. (3) We extensively evaluate and compare the performance of these two heuristics in different scenarios, with different settings, based on a mobility dataset. Our experimental results show that, even in the case where the adversary considers co-locations with only a single friend of the targeted user, the median location privacy of the user is decreased by up to 75% in a typical setting. Even in the case where a user does not disclose any location information, her privacy can decrease by up to 16% due to the information reported by other users. A paramount finding of our work is that users partially lose control over their location privacy as co-locations and individual location information disclosed by other users substantially affect their own location privacy. To the best of our knowledge, this is the first work to quantify the effects of co-location information, that stems from social relationships, on location privacy; thus making a connection between privacy implications of social networks and location privacy.

The remainder of the paper is organized as follows. In Section 2, we define and formalize the system model. In Section 3, we present the optimal localization attack for N users and assess its complexity. In Section 4, we show how this complexity can be reduced by means of approximations. In Section 5, we report on the experimental evaluation of the localization attack with co-locations. In Section 6, we survey the related work. In Section 7, we conclude the paper and suggest directions for the future work.

2 System Model and Formalization

We consider a set of mobile users who move in a given geographical area. While on the go, users make use of some online services to which they communicate potentially obfuscated location (*i.e.,* where they are) and accurate co-location information (*i.e.,* who they are with). We consider that a curious service provider (referred to as the adversary) wants to infer the location of the users from this information, hence tracking them over time. In order to carry out the inference attack, based on which the location privacy of the users is evaluated, the adversary would model the users as described below. Our model is built upon [20] and uses similar notations. Figure 2 gives an overview of the considered scenario.

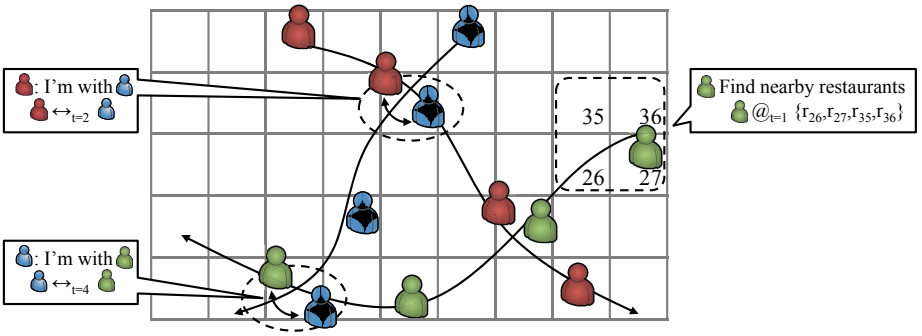


Fig. 2. Scenario of (co-)location exposure. Three users move in a given geographical area. They communicate their potentially obfuscated locations and accurate co-location information to a service provider (*i.e.,* the adversary) who wants to infer their locations.

2.1 Users

We consider a set $\mathcal{U} = \{u_1, \ldots, u_N\}$ of N mobile users who move within a given geographical area that is partitioned into M regions (locations) $\mathcal{R} = \{R_1, \ldots, R_M\}$. Time is discrete and we consider the state of the system (including the locations of the users) at the successive time instants $\{1, \ldots, T\}$. The region in which a user $u \in \mathcal{U}$ is at time instant $t \in \{1, \ldots, T\}$ is called the *actual location* of the user and is denoted by $a_u(t)$. The mobility of the users is modeled by a first order time-homogeneous Markov chain. We denote by $p_u(\rho, r)$ the probability that user u moves from region ρ to region r during one time instant, and by $\pi_u(r)$ the probability that user u is in region r at time t (*i.e.,* the stationary distribution of p_u). We call a co-location the fact that two users are at the same location at some point in time. The fact that users u and v are co-located at time t means that $a_u(t) = a_v(t)$; we denote by $u \leftrightarrow_t v$ the fact that a co-location between users u and v at time t is reported, and we denote by $c_u(t)$ the set of all reported co-locations that involve user u at time t. We define $C_t = \bigcup_{u \in \mathcal{U}} c_u(t)$ and $C = \bigcup_{t=1..T} C_t$.

2.2 Location-Privacy Protection Mechanisms

In order to protect their privacy, users rely on location-privacy protection mechanisms (LPPM) for obfuscating their individual location information before they communicate it to an online service provider. We denote by $u \,@_t\, r'$ the fact that user u reports being at location r' at time t to the online service. The online service observes only the obfuscated location of the users, which we denote by $o_u(t)$ for a user u at time t. We denote by $\mathcal{R}'$ the set of obfuscated locations; typically $\mathcal{R}'$ is the power set of $\mathcal{R}$, as LPPMs can return a set of locations instead of a single one. Typical LPPMs replace the actual location of a user with another location (*i.e.,* adding noise to the actual location) or merge several regions (*i.e.,* reducing the granularity

of the reported location). We model an LPPM by a function that maps a user's actual location to a random variable that takes values in $\mathcal{R}'$, that is, the user's obfuscated location. This means that the locations of a user at different time instants are obfuscated independently from each other and from those of other users. Formally, an LPPM is defined by the function $f_u(r, r')$ which denotes the probability that the LPPM used by u obfuscates location r to r', *i.e.,* $\Pr(o_u(t) = r' \mid a_u(t) = r)$. Let alone the co-location information, our model corresponds to a hidden Markov model (HMM) [1]. We assume that co-location information is not obfuscated and users do not rely on pseudonyms.[1] We denote by $\mathbf{o}(t)$ the vector of the observed locations of all the users at time t. More generally, we use bold notations to denote a vector of values of all users.

2.3 Adversary

The adversary, typically an online service provider (or an external observer who has access to this information, *e.g.,* another user of the social network), has access to the observed locations and co-locations of one or several users and seeks to locate users, at a given time instant, namely carry out a *localization attack*. Because the locations of the users are not independent, given the co-location information, when attacking the location of a given user, the adversary takes into account information potentially about all the users. The attack is performed *a posteriori*, meaning that the adversary has access to the observed traces over the complete period, namely $\{\mathbf{o}(t)\}_{t=1..T}$ and C, at the time of the attack. In addition to the observations during the time period of interest (*i.e.,* $\{1, \ldots, T\}$), the adversary has access to some of the users' past location traces, from which he builds individual mobility profiles for these users, under the form of transition probabilities $\{p_u\}_{u \in \mathcal{U}}$. See [20] for more details about the knowledge construction, in particular on how the mobility profiles can be built from obfuscated traces with missing locations. The mobility profiles constitute, together with the knowledge of the LPPMs used by the users (including their parameters), the adversary's *background knowledge* $\mathcal{K} = \{p_u, f_u\}_{u \in \mathcal{U}}$.

The output of a localization attack that targets a user u at a time instant t, is a *posterior probability distribution* over the set $\mathcal{R}$ of locations.

$$h_t^u(r) \triangleq \Pr(a_u(t) = r \mid \{\mathbf{o}(t)\}_{t=1..T}, C, \mathcal{K}) \quad . \tag{1}$$

2.4 Location Privacy Metric

The location privacy $\mathrm{LP}_u(t)$ of a user u at time t, with respect to a given adversary, is captured by the expected error of the adversary when performing a localization attack [20]. Given the output $h_t^u(\cdot)$ of the localization attack, the location privacy writes:

$$\mathrm{LP}_u(t) \triangleq \sum_{r \in \mathcal{R}} h_t^u(r) \cdot d(r, a_u(t)) \quad , \tag{2}$$

[1] Note that even if pseudonyms are used, the identity of the users can be inferred by using their social network [18] or their locations [20].

where $d(\cdot\,,\cdot)$ denotes a distance function on the set $\mathcal{R}$ of regions, typically the Haversine distance between the centers of the two regions.

3 Optimal Localization Attack

Without co-location information (as in [20]) and under the assumptions described in the previous section, the localization problem translates to solving a HMM inference problem, for which the *forward-backward* algorithm is a known solution. Essentially, the forward-backward algorithm defines forward and backward variables that take into account the observations before and after time t, respectively. The forward variable is the joint probability of location of user at time t and all the observations up to, and including, time t. The backward variable is the conditional probability of all observations after time t, given the actual location of user at that time instant. Then, the posterior probability distribution of the possible locations for the targeted user is obtained by combining (*i.e.,* multiplying and normalizing) the forward and backward variables. With co-location information, the locations of the users are not mutually independent: as soon as two users are co-located at some point in time t, their locations, before and after time t, become dependent. Actually, the fact that two users meet a same third user (even if they meet her at different time instants) suffices to create some dependencies between their locations; this means that, to perform the localization attack on a user, the adversary must take into account the locations (*i.e.,* the obfuscated location information and the co-location information) of all the users who are connected to u by a *chain* of co-location (*i.e.,* the connected component of u in the co-location graph). Formally speaking, it means that the adversary cannot rely only on the *marginal* distributions of the users' location; instead he must consider the *joint* distributions. In other words, co-locations turn N disjoint inference problems (*i.e.,* HMM problems solved by the forward-backward algorithm) into a joint inference problem.

To solve the localization problem, we consider the users jointly; we show that it translates to an HMM problem that we solve using a forward-backward algorithm. For a set $\mathcal{U}$ of users and a time t, we define the following forward and backward variables:

$$\alpha_t^{\mathcal{U}}(\mathbf{r}) \triangleq \Pr\left(\mathbf{o}(1)\ldots,\mathbf{o}(t), C_1,\ldots,C_t, \mathbf{a}(t)=\mathbf{r}\,|\,\mathcal{K}\right) \tag{3}$$

$$\beta_t^{\mathcal{U}}(\mathbf{r}) \triangleq \Pr\left(\mathbf{o}(t+1)\ldots,\mathbf{o}(T), C_{t+1},\ldots,C_T\,|\,\mathbf{a}(t)=\mathbf{r},\mathcal{K}\right)\ , \tag{4}$$

where $\mathbf{r}$ denotes a vector of size N,*i.e.,* $\mathbf{r} \in \mathcal{R}^N$, and represents the actual locations of all users at a single time instant. These variables can be defined recursively (over t) and, unlike in the case where no co-location observations are available, their expressions involve the co-location information. More specifically, it can be proved that for all $\mathbf{r} \in \mathcal{R}^N$, we have[2]

[2] For the sake of simplicity and clarity, we define the variables at $t = 0$ even though no observations are made at this time instant.

$$\alpha_t^{\mathcal{U}}(\mathbf{r}) = \begin{cases} \pi_{\mathcal{U}}(\mathbf{r}) & \text{if } t = 0 \\ \dfrac{\mathbf{1}_t(\mathbf{r}, C)}{\sum_{C'} \mathbf{1}_t(\mathbf{r}, C')} \cdot f_{\mathcal{U}}\left(\mathbf{r}, \mathbf{o}(t)\right) \displaystyle\sum_{\boldsymbol{\rho} \in \mathcal{R}^N} \alpha_{t-1}^{\mathcal{U}}(\boldsymbol{\rho}) \cdot p_{\mathcal{U}}\left(\boldsymbol{\rho}, \mathbf{r}\right) & \text{if } t > 0 \end{cases} \tag{5}$$

and

$$\beta_t^{\mathcal{U}}(\mathbf{r}) = \begin{cases} \displaystyle\sum_{\boldsymbol{\rho} \in \mathcal{R}^N} \frac{\mathbf{1}_t(\boldsymbol{\rho}, C)}{\sum_{C'} \mathbf{1}_t(\boldsymbol{\rho}, C')} \cdot \beta_{t+1}^{\mathcal{U}}(\boldsymbol{\rho}) \cdot p_{\mathcal{U}}\left(\mathbf{r}, \boldsymbol{\rho}\right) \cdot f_{\mathcal{U}}\left(\boldsymbol{\rho}, \mathbf{o}(t+1)\right) & \text{if } t < T \\ 1 & \text{if } t = T \end{cases} \tag{6}$$

where for $\mathbf{r} = (r_1, \ldots, r_N) \in \mathcal{R}^N$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_N) \in \mathcal{R}^N$, $\mathbf{r}' = (r'_1, \ldots, r'_N) \in \mathcal{R}'^N$, $\pi_{\mathcal{U}}(\mathbf{r}) = \prod_{i=1}^N \pi_{u_i}(r_i)$, $f_{\mathcal{U}}(\mathbf{r}, \mathbf{r}') = \prod_{i=1}^N f_{u_i}\left(r_i, r'_i\right)$, $p_{\mathcal{U}}\left(\boldsymbol{\rho}, \mathbf{r}\right) = \prod_{i=1}^N p_{u_i}(\rho_i, r_i)$, and $\mathbf{1}(\cdot\,, \cdot)$ is the indicator function that returns 1 if the locations of the users are *consistent* with the co-location information reported at time t, and 0 otherwise. That is, formally,

$$\mathbf{1}_t(\mathbf{r}, C) = \begin{cases} 1 & \text{if } \forall (u_i \leftrightarrow_t u_j) \in C_t,\ r_i = r_j \\ 0 & \text{otherwise} \end{cases} . \tag{7}$$

In other words, the indicator function captures whether the users for which a co-location was reported are indeed at the same locations in $\mathbf{r}$. As the adversary has no knowledge about the way co-locations are reported, the distribution of the sets of reported co-locations, given the actual locations of the users, is modeled with a uniform distribution.

The intuition behind Equation (5) is that the forward variable at time t can be expressed recursively, with respect to time, by combining, for all possible locations of the users at time $t-1$: (1) the joint probability that the users were at location $\boldsymbol{\rho}$ at time $t-1$ and reported the obfuscated locations observed by the adversary up to time $t-1$ (this is captured by $\alpha_{t-1}^{\mathcal{U}}$), (2) the joint probability that the users move from the locations $\boldsymbol{\rho}$ to the locations $\mathbf{r}$ (this is captured by $p_{\mathcal{U}}$), (3) the joint probability that the users obfuscate their locations $\mathbf{r}$ to that observed by the adversary $\mathbf{o}(t)$ (this is captured by $f_{\mathcal{U}}$) and that the locations $\mathbf{r}$ of the users are consistent with the co-locations reported at time t. Because users obfuscate their locations independently from each other, the joint obfuscation probability is the product of the individual obfuscation probabilities (hence the expression of $f_{\mathcal{U}}$). The same applies to $p_{\mathcal{U}}$. The same intuition lies behind Equation (6).

The indicator function $\mathbf{1}_t(\cdot\,, \cdot)$ accounts for the co-location information in the localization attack by ruling out the *impossible* (*i.e.,* inconsistent with the reported co-locations) user locations, hence further narrowing down the set of possible locations for the users involved in a co-location. Schematically speaking (with a deterministic vision, for the sake of clarity), the set of possible locations for a user u_i (at time t), co-located with a user u_j, consists of the locations that can be obfuscated into the location reported by u_i at time t and that can be reached (according to u_i's mobility profile) from a possible location of u_i at time

$t-1$ **and** that can be obfuscated into the location reported by u_j at time t **and** that can be reached (according to u_j's mobility profile) from a possible location of u_j at time $t-1$.

Finally, the posterior probability distribution of the users' locations can be computed based on the forward and backward variables, by using the following formula, for $u_i \in \mathcal{U}$ and at time t:

$$h_t^{u_i}(r) = \Pr\left(a_{u_i}(t) = r \mid \{\mathbf{o}(t)\}_{t=1..T}, C, \mathcal{K}\right) = \frac{\sum_{\mathbf{r}\in\mathcal{R}^N \mid r_i=r} \alpha_t^{\mathcal{U}}(\mathbf{r}) \cdot \beta_t^{\mathcal{U}}(\mathbf{r})}{\sum_{\mathbf{r}\in\mathcal{R}^N} \alpha_t^{\mathcal{U}}(\mathbf{r}) \cdot \beta_t^{\mathcal{U}}(\mathbf{r})} \ . \quad (8)$$

We now evaluate the complexity of the joint localization attack. The first observation is that the size of the state space (*i.e.,* the locations of all users) is M^N. To attack a user at time t, the adversary needs to compute the values of α *up to* time t and the values of beta *down to* time t.[3] At each time instant, the adversary needs to compute the values of these two variables for all possible values of their inputs $\mathbf{r} \in \mathcal{R}^N$ (there are M^N possible values for $\mathbf{r}$). The computation of each of these values requires summing over the M^N possible locations $\boldsymbol{\rho}$ at time $t-1$; for each of the possible locations, the computation of one element of the sum takes $\Theta(N)$ operations. Therefore, the computation of the forward and backward variables, at all time instants, for all possible values of the localizations is $\Theta(NTM^{2N})$ operations. Note that the complexity is the same whether the adversary attacks one or all the users at one or all time instants. In fact, the adversary can pre-compute the h_t^u for all u and all t with a complexity that is dominated by that of the computations of the forward and backward variables. In summary, the complexity of the localization attack on one or all of the users in $\mathcal{U}$ is

$$c_{\text{opt}}(N, T, M) = \Theta(NTM^{2N}) \ . \quad (9)$$

The complexity of the optimal localization attack is prohibitively high and prevents its use for the entire set of users of a mobile social network; the optimal localization attack is tractable only for small values of N, *i.e.,* 2 and 3. In the next section, we propose heuristics for performing low-complexity approximate localization attacks.

4 Approximate Localization Attack

We propose two low-complexity heuristics for performing approximate localization attacks. Essentially, the first selects a small set of users to consider when

[3] The best way to do this is to use dynamic programming, *i.e.,* compute the α_t (and storing its values) iteratively for increasing t and compute the β_t (and store the values) iteratively for decreasing t.

attacking a target user and performs an optimal joint localization attack on this small set of users (*i.e.,* considering only the co-locations between these users). The intuition behind this heuristic is that the locations of a user are significantly correlated with those of only a limited number of users (*e.g.,* a few co-workers during work hours, and her family and close friends the rest of the time). The second makes use of individual forward-backward variables (one for each user of the entire set of users) and computes their values at each time instant, based on the considered user's individual variable at time $t-1$ and the reported locations of the users co-located with her at time t, hence disregarding the dependencies stemming from past co-locations. The intuition behind this heuristic is that the dependency between two users' locations fades relatively quickly over time after they meet.

4.1 Heuristic 1: Limited User Set Approximation

As discussed in Section 3, the optimal localization attack can be efficiently performed only on small sets of users. This is because location of a target user u depends on locations of *all* other users that are connected to u in the co-location graph (where there is an edge between two users u and v if $u \leftrightarrow_t v$ for some time t). The rationale of our first approximation is to limit the number of users, on which the target user's location depends, and to consider only those that have high location correlation with u. Concretely, we choose the user(s) that have the largest number of reported co-locations with the targeted user and we perform an optimal localization attack on the resulting set of users. We call these users the *co-targets* of the targeted user. Depending on his computational power, the adversary can choose one or two such users (*i.e.,* $N = 2$ or $N = 3$) to attack the target with. The co-targets of a user u are chosen as follows:

$$\text{co-target}_1(u) \triangleq \underset{v \in \mathcal{U} \setminus \{u\}}{\operatorname{argmax}} |\{t \in \{1, \ldots, T\} \mid u \leftrightarrow_t v\}| \tag{10}$$

$$\text{co-target}_2(u) \triangleq \underset{v \in \mathcal{U} \setminus \{u, u'\}}{\operatorname{argmax}} |\{t \in \{1, \ldots, T\} \mid u \leftrightarrow_t v\}| + |\{t \in \{1, \ldots, T\} \mid u' \leftrightarrow_t v\}| \tag{11}$$

where $u' = \text{co-target}_1(u)$ and $|\cdot|$ denotes the cardinality of the set. More specifically, the first co-target of a user u is the user with whom u has the more reported co-locations during the time interval considered for the localization attack. The second co-target of u is chosen so as to maximize the number of co-locations with u **plus** the number of co-locations with u's first co-target. Note that the set of considered users can be different for every targeted user; in particular $v = \text{co-target}_1(u) \not\Longrightarrow u = \text{co-target}_1(v)$. The complexity of this heuristic is $\Theta(TM^4)$ for $N = 2$ and $\Theta(TM^6)$ for $N = 3$ (obtained by replacing N by its value in the generic expression (9) of the complexity of the optimal attack).

4.2 Heuristic 2: Independence Approximation

As discussed in Section 3, the need to jointly consider the locations of all the users, which cause the explosion of the state space size and thus the high

complexity of the attack, stems from the fact that their locations are not independent as soon as co-locations are reported. The rationale behind our second heuristic is to ignore the mobility profiles of the co-located users, hence alleviating the need to take into account their past locations, which causes the state space explosion) and to consider only their reported co-locations to improve the inference of the target user's location at the considered time instant. This comes down to considering the locations reported by the users co-located with u, as if u had reported these obfuscated locations herself (as depicted in Figure 1a). We define individual forward and backward variables for each user and we couple them upon co-locations, as follows:

$$\hat{\alpha}_t^u(r) \triangleq \begin{cases} \pi_u(r) & \text{if } t = 0 \\ \prod\limits_{u' \mid u \leftrightarrow_t u'} f_{u'}\left(r, o_{u'}(t)\right) \cdot f_u\left(r, o_u(t)\right) \cdot \sum\limits_{\rho \in \mathcal{R}} \hat{\alpha}_{t-1}^u(\rho)\, p_u(\rho, r) & \text{otherwise} \end{cases} \tag{12}$$

and

$$\hat{\beta}_t^u(r) \triangleq \begin{cases} 1 & \text{if } t = T \\ \sum\limits_{\rho \in \mathcal{R}} \hat{\beta}_{t+1}^u(\rho)\, p_u(r, \rho)\, f_u(\rho, o_u(t+1)) \prod\limits_{u' \mid u \leftrightarrow_{t+1} u'} f_{u'}(\rho, o_{u'}(t+1)) & \text{otherwise} \end{cases} \tag{13}$$

Finally, when performing a localization attack on user u, the posterior distributions of the locations of the users co-located with u at time t are taken into account. More specifically, we estimate the probability distribution of user u's location at time t by

$$\hat{h}_t^u(r) \triangleq \frac{\hat{\alpha}_t^u(r) \cdot \hat{\beta}_t^u(r) \cdot \prod\limits_{u' \mid u \leftrightarrow_t u'} \hat{\alpha}_t^{u'}(r)\ \hat{\beta}_t^{u'}(r)}{\sum\limits_{r' \in \mathcal{R}} \left(\hat{\alpha}_t^u(r')\ \hat{\beta}_t^u(r') \prod\limits_{u' \mid u \leftrightarrow_t u'} \hat{\alpha}_t^{u'}(r')\ \hat{\beta}_t^{u'}(r') \right)} . \tag{14}$$

We now compute the complexity of this heuristic. To perform a localization attack on a user, the adversary needs to compute the individual variables of all the users that are connected to the target by a chain of co-location, that is N users at most. The computation of a value $\hat{\alpha}$ and $\hat{\beta}$ (for a given t and a given r), in the worst case (*i.e.,* when all the users are co-located), takes $\Theta(NM)$ operations; and TM such values need be computed for each user. Therefore, the complexity of this heuristic is $\Theta(N^2TM^2)$.

5 Experimental Evaluation

We evaluate the effect of co-locations on users' location privacy, with respect to the various localization attacks presented in the previous sections, by using a dataset of real mobility traces.

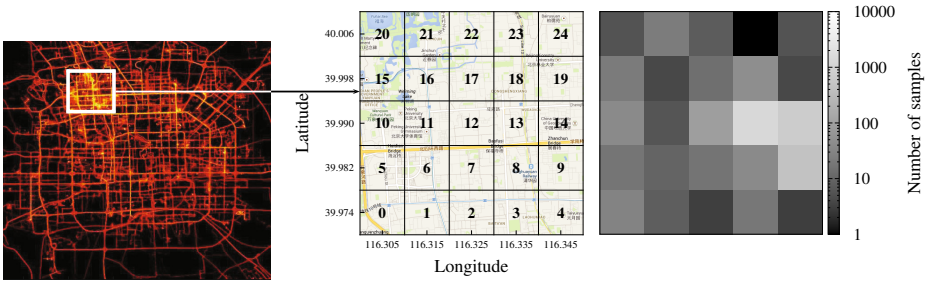


Fig. 3. Illustration of the dataset used in the evaluation. Most traces are located in the region of Beijing (left); we focus on a small active area that corresponds to the campus of the Tsinghua University and we partition it by using a 5×5 square grid (middle). The heat-map (right) shows the number of samples in each region (logscale).

5.1 Dataset, Methodology and Experimental Setup

The dataset was collected by Microsoft Research Asia, in the framework of the GeoLife project [24]. It comprises the GPS traces (*i.e.,* sequences of time-stamped latitude-longitude couples, sampled at a rate of one point every 1-5 seconds) of 182 users, collected over a period of over three years. The GPS traces are scattered all over the world; but most of them are located in the region of Beijing, China. We processed the data as follows, in order to fit in our formalism.

Space Discretization. We select the area of $\sim$4.4 km $\times$ 4.4 km, within Beijing, that contains the largest number of GPS samples, and we filter out GPS samples that are outside of this area. This geographic area corresponds to the campus of the Tsinghua University (longitude ranging from 116.3 to 116.35 and latitude ranging from 39.97 to 40.01, see Figure 3). We partition the selected area into 25 regions by using a 5×5 square grid. The GPS coordinates of each sample are translated into the region (*i.e.,* the grid cell) they fall into.

Time Discretization. We split the continuous time interval into one-hour time sub-intervals, which correspond to time instants in our formalism. For each time sub-interval t and for each user u, we set the user's actual location in that time interval (*i.e.,* $a_u(t)$) to the region corresponding to the sample that is the closest to the midpoint of the considered time sub-interval. If a user's trace does not contain any sample in a given time sub-interval, the user's actual location is set to a dummy region $r_\perp$, leaving us with partial user traces.

Co-location Generation. As the dataset does not contain explicit co-location information reported by the users, we use synthetic co-locations that we generate as follows: At each time instant, we generate a co-location between two users if their discretized actual locations are the same (and different from $r_\perp$). Because in real-life not all such situations correspond to actual co-location and because even actual co-locations are not necessarily reported, in our evaluation we take

into account only a proportion ω (ranging from 0% to 100%) of the synthetic co-locations.

For each user, we compute the number of co-locations she has with every other user in the dataset, across the full user traces. We keep only the users for which there exists another user with whom they have at least 200 co-locations. For these users, we consider their *common* time interval (*i.e.,* the longest time interval during which all these users have at least one sample); we obtained an interval of ∼6000 hours. Within the common interval, we sample 10 short traces of 300 continuous hours such that (1) all users have at least 10% of valid samples (*i.e.,* , different from $r_\perp$) and (2) all users have at least 20 co-locations with their co-target$_1$ (as defined in Equation (11)). This leaves us with a total of 5 users.

User Mobility Profiles Construction. We build the mobility profiles $\{p_u\}_{u\in\mathcal{U}}$ of the users based on their entire discretized traces by counting the transitions from any region to any region (in $\mathcal{R}$) in one time instant.

Obfuscation. We consider that users report a single (or none), potentially obfuscated, location at each time instant.[4] This means that the set $\mathcal{R}'$ in which the obfuscated location $o_u(\cdot)$ takes values is $\mathcal{R}\cup\{r_\perp\}$. We consider, for each user u, that two location-privacy protection mechanisms are used together: First, the location is hidden (*i.e.,* obfuscated to $r_\perp$) with a probability λ_u and then, if the location has not been hidden, it is replaced by a region (chosen uniformly at random) at a distance of at most d_u from the user's actual discretized location (*i.e.,* a region). If the actual location of a user is not known (*i.e.,* set to $r_\perp$), the LPPM returns $r_\perp$ with probability 1. In our evaluation, we vary λ_u from 0 to 1 and we set d_u to the size of one grid cell; this means that, if it is not hidden, a user's location is obfuscated either to its actual value (with probability 0.2) or to one of the four adjacent regions (*e.g.,* 2, 6, 8 and 12 for region 7 in Figure 3), each with probability 0.2.

Privacy Evaluation. We evaluate the location privacy of the users, and the effect of co-locations on it, based on the metric defined in (2). For each user and for each short trace, we generate 20 random obfuscated traces (remember that obfuscation is a random process) and we perform a localization attack on each of them. We compute the average location privacy of each user across the different obfuscated traces and across the different time instants. Time instants for which the location of a user is not known (*i.e.,* set to $r_\perp$) are not taken into account in the computation of their average over time.

Limitations. Due to the synthetic nature of the reported location and co-location information in our data source, our experimental setup does not perfectly reflect on a real usage case. Therefore, the results presented in this section cannot directly be interpreted as the magnitude of the threat in real-life. Yet, we believe that it suffices to get insight into the effect of co-locations on location

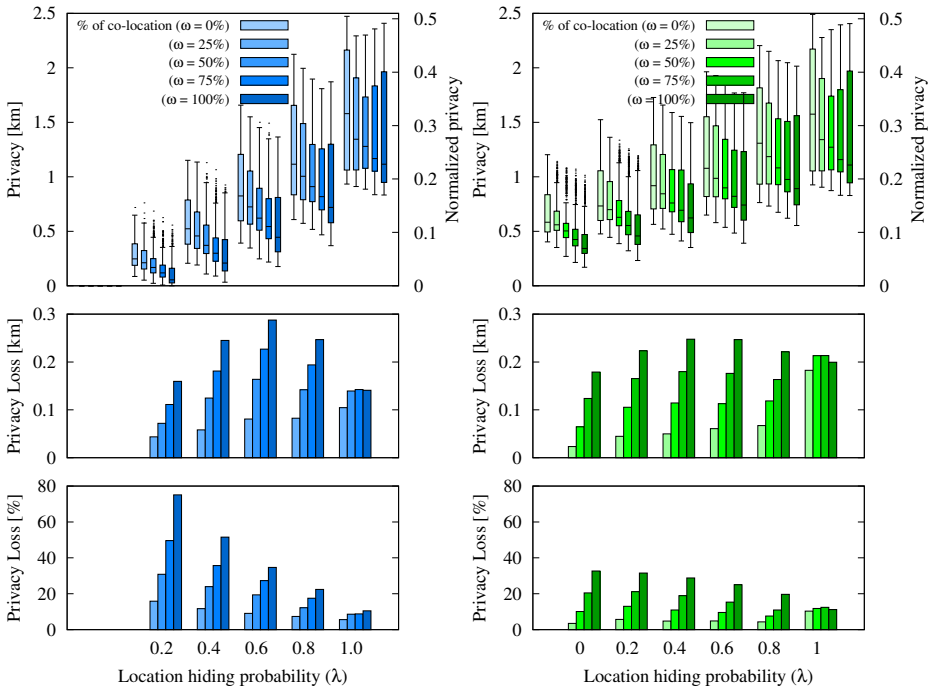
[4] We make this assumption because of the limited size of the considered grid and we leave the case where LPPMs output a *set* of locations to future work.

privacy, the sources of privacy loss, and the relative performance of the proposed heuristics. Also, the number of users considered in our evaluation (*i.e.,* 5) is relatively small. Hence, the results may not be representative of the entire population. In order to overcome the aforementioned shortcomings, we intend to collect a large-scale dataset from an existing social network. We also intend to run experiments on large grids (*i.e.,* larger than the 5×5 grid used in the evaluation).

5.2 Experimental Results

We now experimentally evaluate the algorithms, presented in Section 4, in different scenarios, with different settings. The goal of our evaluation is to assess the raw performance of our heuristics, but also to compare them. In addition, we analyze the effect of the different parameters of the model (including the individual LPPM settings of the users and the *differences* between the individual LPPM settings of the users) and of the set of co-locations considered in the localization attack.

Effects of Co-locations and LPPM Settings. We begin our evaluation by analyzing the effect of (1) the proportion ω of reported co-location and (2) the LPPM settings (*i.e.,* w/ or w/o obfuscation and the location hiding probability λ, assumed to be the same across users) in the case of two users, *i.e.,* the target user and her first co-target are considered jointly in an optimal localization attack, namely the limited user set approximation with $N = 2$. The results are depicted in Figure 4. The left sub-figure shows the case where no obfuscation is used (*i.e.,* the users disclose their *actual* locations with probability $1-\lambda$ and hide them completely otherwise), whereas the right sub-figure shows the case where obfuscation is used (*i.e.,* the users disclose their *obfuscated* locations, specifically a region chosen uniformly at random among the actual location and the four immediate neighboring regions, with probability $1 - \lambda$ and hide them otherwise). The top graphs show a box-plot representation (*i.e.,* first quartile, median, third quartile and outliers) of the users' location privacy expressed in terms of the expected error of the adversary, in kilometers (left axis) and in proportion of the size of the considered geographic area (right axis). For each couple of values (λ, ω), we draw one box-plot to aggregate the data-points obtained for all users and for all the 20 randomly generated obfuscated versions of each of the considered actual trace. Note that without obfuscation, the case $\lambda = 0$ leads to zero privacy as users *always* disclose their *actual* locations. It can be observed that the proportion of reported co-locations consistently decreases the location privacy of the users. To quantify this decrease, we plot (middle and bottom graphs) the privacy loss caused by the use of co-location information, with respect to the case where co-locations are ignored (or not available), *i.e.,* $\omega = 0\%$. We show both the median absolute privacy loss (in kilometers, middle graph) and the median relative privacy loss (in percentage of the privacy in the case $\omega = 0\%$, bottom graph). Note that the median privacy loss is **not** equal to the difference of the median privacy. Consider for example, the case $\lambda = 0.4$ and $\omega = 50\%$:

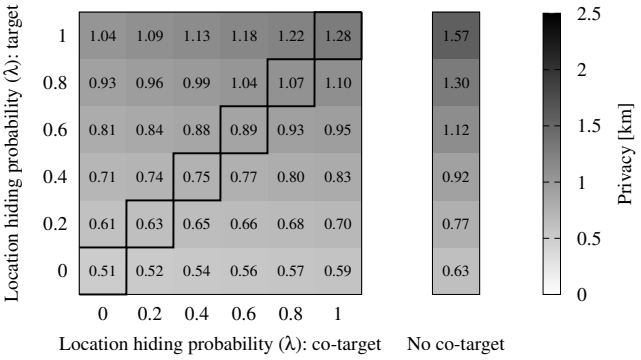


(a) Without obfuscation (b) With obfuscation

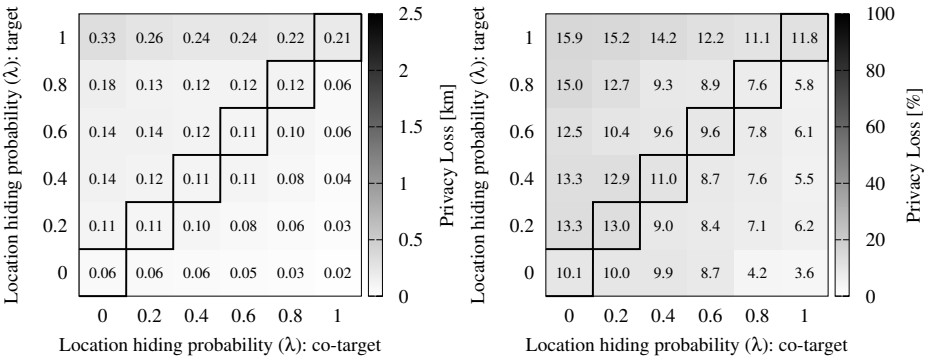
Fig. 4. Privacy (top), absolute privacy loss (middle) and relative privacy loss (bottom) for the limited user set attack with $N = 2$ users. The privacy *loss* is expressed wrt the case where no co-locations are reported ($\omega = 0\%$); the histograms show median values. Co-location information decreases privacy. The relative privacy loss is higher for small values of the hiding probability and without obfuscation.

in the case without obfuscation the median privacy loss is approximately 125m, which corresponds to a decrease of 25%. The median absolute privacy loss can go up to 290m ($\lambda = 0.6$, $\omega = 100\%$) and the median relative privacy loss up to 75% ($\lambda = 0.2$ and $\omega = 100\%$). We observe the same trend, with a more modest loss, in the case where obfuscation is used. For the rest of the evaluation, we focus on the case where users do obfuscate their locations and report $\omega = 50\%$ of the co-locations.

Effects of the differences of Individual LPPM Settings. Here, we analyze the effect of the differences, in the users' LPPM settings, on the location privacy (loss) due to co-locations. To do so, we focus on the case of two users, a target and her co-target, both who obfuscate their location but with different hiding probabilities λ_{target} and $\lambda_{\text{co-target}}$. We perform a joint optimal localization attack. The results are depicted in Figure 5 under the form of heat-maps that represent the target user's location privacy (a) as well as her absolute (b) and relative (c)



(a) Privacy



(b) Absolute privacy loss (wrt to $\omega = 0\%$) (c) Relative privacy loss (wrt to $\omega = 0\%$)

Fig. 5. Median values of the target's location privacy (loss), for the limited user set attack with $N = 2$ users, when the target and her co-target have different values of λ (with obfuscation and $\omega = 50\%$). The diagonals correspond to the values of Figure 4b.

privacy loss (wrt the case $\omega = 0\%$) as functions of the respective LPPM settings $\lambda_{\text{co-target}}$ (x-axis) and λ_{target} (y-axis).

A first observation is that co-locations always decrease the privacy of the target (*i.e.,* all values in Figure 5b are positive) and that the more information the co-target discloses, the worse the privacy of the target is (*i.e.,* the cells of the heat-map depicted in Figure 5a become lighter, when going from right to left on a given row).

The diagonals of the heat-maps correspond to the case $\lambda_{\text{co-target}} = \lambda_{\text{target}}$, which is depicted in more details in Figure 4. The region of the heat-map above the diagonal corresponds to the case where the target is more *conservative*, in terms of her privacy attitude, than her co-target (*i.e.,* $\lambda_{\text{co-target}} < \lambda_{\text{target}}$). It can be observed that the information disclosed by the target herself compromises her privacy more than the information disclosed by her co-target, *e.g.,* the cell (0.6,0) is lighter (which means that the target's privacy is lower) than the cell (0,0.6).

By comparing the columns "$\lambda_{\text{co-target}} = 1$" and "no co-target" (two right-most columns in Figure 5a), we can observe the privacy loss stemming from the use, through the co-location information, of the co-target's mobility profile alone (as the co-target never discloses her location). This is substantial.

Finally, in the extreme case where the target never discloses location information and her co-target always does (top-left cell of the heat-maps in Figures 5b and 5c), the privacy loss for the target is 330m, which corresponds to a decrease of 16%. This case (and in general the cases where the target never discloses location information, *i.e.,* the top row of the heat-maps) highlights the fact that, as reported co-locations involve two users, users lose some control over their privacy: Without revealing any information about herself, a user can still have her privacy decreased by other users, due to co-location information.

For the rest of the evaluation, we focus on the case where all users have the same LPPM settings (*i.e.,* same values of λ).

Comparison of the Proposed Heuristics. Here, we compare, through experimentation (we leave the analytical comparison to future work), the proposed inference algorithms for the localization attack, by taking into account different scenarios, as depicted in Figure 6. In scenario (a), we consider, in turn, all target users in our set and perform an individual localization attack on each of them, using only their own reported locations and no co-locations. This corresponds to the baseline case $\omega = 0\%$, which was presented in detail in Figure 4b. We then consider the case of an adversary that exploits co-locations. We assume the adversary observes a limited proportion, $\omega = 50\%$, of the existing co-locations. Scenario (b) corresponds to the case of an adversary that, in order to attack a target user, performs an optimal joint inference attack on the target and her co-target, as described in Section 3. This scenario corresponds to the case $\omega = 50\%$ in Figure 4b. Scenarios (c) and (d) correspond to the case of an adversary that performs an optimal joint attack on the target and her **two co-targets**. We distinguish two cases, (c) – in which the co-locations between the co-targets are ignored and (d) – in which all co-locations between any of the three users are considered. We make this distinction solely to quantify the privacy loss stemming from the use of co-locations that do not directly involve the target. In practice, an adversary would always consider scenario (d) as it takes into account more information at no extra cost. Finally we consider scenario (e), that corresponds to an adversary that uses reported all co-locations but solves an individual inference problem for each user, as described in 4.2.

Figure 7 shows the results of our comparison. The graph on the left shows a box-plot representation of users' privacy, for each of scenarios (a)-(e). To quantify the different effects on the users' privacy of the set of considered co-locations and of the heuristic used, we show (right) the absolute and relative privacy loss, with respect to scenario (a), for each of the scenarios (b)-(e). It can be observed by comparing scenarios (a)-(d) that, unsurprisingly, the users' privacy decreases with the amount of considered co-locations. However, the comparison between scenarios (c) and (d) shows that co-locations between the target's co-targets does not significantly improve the performance of the localization attack. Finally,

we observe that the second heuristic, which takes into account all co-locations outperforms the first heuristic ($N \leq 3$), at a lower computational cost.

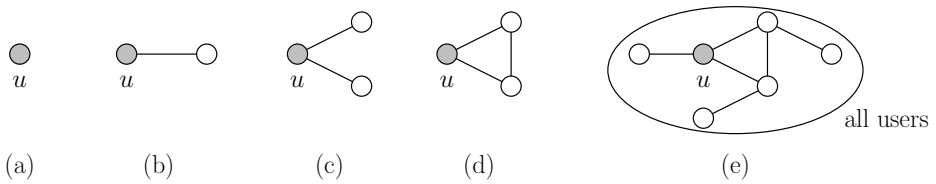


Fig. 6. Co-locations considered in the evaluation: (a) no co-locations, (b) only co-locations between the target and co-target$_1$ (Heuristic 1, $N = 2$), (c) only co-locations between the target and co-target$_1$ and between the target and co-target$_2$ (Heuristic 1, $N = 3$), (d) all co-locations between the target, co-target$_1$ and co-target$_2$ (Heuristic 1, $N = 3$), (e) all co-locations (Heuristic 2). In scenarios (b)-(e), we consider that $\omega = 50\%$ of the co-locations are reported.

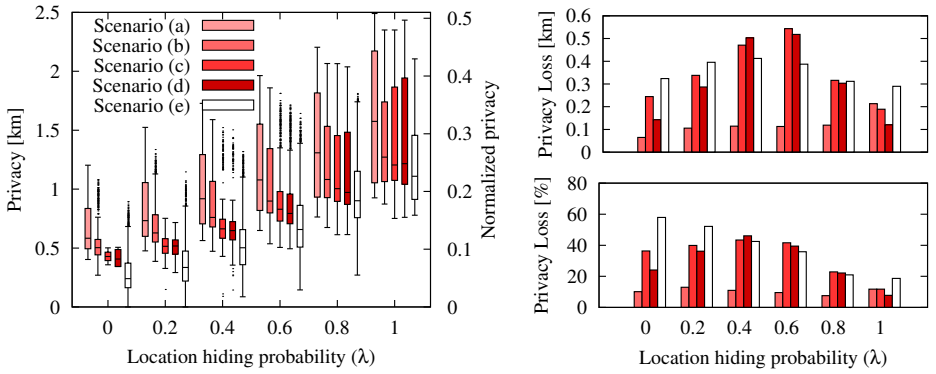


Fig. 7. Comparison of the different localization attacks for the scenarios (a)-(e) depicted in Figure 6. The privacy loss (right) is evaluated wrt scenario (a).

6 Related Work

Location is identity. Even if the set of locations shared by a user is anonymized, and her true identity is hidden from the location-based service provider, the observed trajectories can be re-identified [5, 9, 12, 15]. This attack is made by linking available information about users' mobility in the past with their observed traces. To protect against such attacks, many location obfuscation mechanisms have been proposed in the literature; they suggest users hide their locations at certain locations, or reduce the accuracy or granularity of their reported locations [4, 8, 13]. These techniques increase users' privacy by making it more difficult for an adversary to de-anonymize users and localize or track them over time. The location privacy of users in such settings can be computed using the

expected error of an adversary in estimating their true locations [20]. In such an inference framework, an adversary has a background knowledge on users' mobility models; this is used to reconstruct the full trajectories of the users, given the anonymized and obfuscated observed traces.

The adversary's information, however, is not limited to mobility models. With most users being members of social networks, an adversary can de-anonymize location traces by matching the graph of co-traveler users with their social network graph [21]. Co-travelers are those who have been in each others' physical proximity for a considerable number of times. Researchers have extensively studied the problem of inferring social ties between users based on their physical proximity [3,7]. Recent revelations about NSA surveillance programs also show that this type of information is of great use for tracking and identifying individuals [2].

The correlation between different users' information also opens the door to a new type of privacy threat. Even if a user does not reveal much information about herself, her privacy can be compromised by others. In [11], the authors study how information revealed, from pictures, by a user's friends in social networks can be used to infer private information about her location. Private information about, for example, user profile and her age can also be inferred from shared information on online social networks [6,17]. Mobile users, connecting to location-based services from a same IP address, can also compromise the privacy of those who want to keep their location private [23]. The loss in privacy, due to other users, has also been shown in other contexts such as genomics [10, 14].

Extracting co-location information about users, i.e., who is with whom, is becoming increasingly easier. More specifically, with the proliferation of mobile social networks, where users can check-in themselves and others to different locations, the threat of available co-location information on users' location privacy is clear (as pointed out in [22]). Despite the mentioned works on quantifying the location privacy and the privacy of users in social networks, as well as the extensive research on privacy loss due to others, there has not been a study on evaluating location privacy considering co-location information. We bridge the gap between studies on location privacy and social networks, and we propose the first analytical framework to quantify the effects of co-location information on location privacy, where users can also make use of obfuscation mechanisms.

7 Conclusion

In this paper, we have studied the effect on users' location privacy when co-location information is available, in addition to individual (obfuscated) location information. To the best of our knowledge, this is the first paper to quantify the effects of co-location information, that stems from social relationships between users, on location privacy; as such it constitutes a first step towards bridging the gap between studies on location privacy and social networks. We have shown that, by considering the users' locations jointly, an adversary can exploit co-location information to better localize users, hence decreasing their individual privacy. Although the optimal joint localization attack has a prohibitively

high computational complexity, the polynomial-time approximate inference algorithms that we propose in the paper provide good localization performance. An important observation from our work is that a user's location privacy is no longer entirely in her control, as the co-locations and the individual location information disclosed by other users significantly affect her own location privacy.

The message of this work is that protection mechanisms must not ignore the social aspects of location information. Because it is not desirable to report dummy lists of co-located users (as this information is displayed on the users' profiles on social networks), a location-privacy preserving mechanism needs instead to generalize information about co-located users (*i.e.,* replace the names of the co-located users by the type of social tie, *e.g.,* "with two friends") or to generalize the time (*i.e.,* replace the exact time of the co-location with the period of the day, *e.g.,* replacing 11am with "morning", when the co-location is declared *a posteriori*) of a social gathering as well as the locations of users at other locations, in order to reduce the effectiveness of the attacks we suggested in this paper. We intend to tackle the design of social-aware location-privacy protection mechanisms (running on the users' mobile devices) to help the users assess and protect their location privacy when co-location information is available.

Acknowledgments. The authors are thankful to Stefan Mihaila for his help in collecting useful statistics about the use of co-location on Foursquare. This work was partially funded by the Swiss National Science Foundation with grant 200021-138089.

References

1. Baum, L.E., Petrie, T.: Statistical inference for probabilistic functions of finite state markov chains. The Annals of Mathematical Statistics 37(6), 1554–1563 (1966)
2. How the NSA is tracking people right now (2013), `http://apps.washingtonpost.com/g/page/national/how-the-nsa-is-tracking-people-right-now/634/` (last visited: February 2014)
3. Crandall, D.J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J.: Inferring social ties from geographic coincidences. Proc. of the National Academy of Sciences (PNAS), 1–6 (2010)
4. Damiani, M.L., Bertino, E., Silvestri, C.: The PROBE framework for the personalized cloaking of private locations. Transactions on Data Privacy 3, 123–148 (2010)
5. De Mulder, Y., Danezis, G., Batina, L., Preneel, B.: Identification via location-profiling in GSM networks. In: WPES 2008: Proc. of the 7th ACM Workshop on Privacy in the Electronic Society, pp. 23–32 (2008)
6. Dey, R., Tang, C., Ross, K., Saxena, N.: Estimating age privacy leakage in online social networks. In: INFOCOM 2012: Proc. of the 31st Annual IEEE Int'l Conf. on Computer Communications, pp. 2836–2840 (2012)
7. Eagle, N., Pentland, A., Lazer, D.: Inferring Friendship Network Structure by Using Mobile Phone Data. Proc. of the National Academy of Sciences (PNAS) 106, 15274–15278 (2009)
8. Ghinita, G., Damiani, M.L., Silvestri, C., Bertino, E.: Preventing velocity-based linkage attacks in location-aware applications. In: GIS 2009: Proc. of the 17th ACM Int'l Symp. on Advances in Geographic Information Systems, pp. 246–255 (2009)

9. Golle, P., Partridge, K.: On the anonymity of home/work location pairs. In: Tokuda, H., Beigl, M., Friday, A., Brush, A.J.B., Tobe, Y. (eds.) Pervasive 2009. LNCS, vol. 5538, pp. 390–397. Springer, Heidelberg (2009)
10. Gymrek, M., McGuire, A.L., Golan, D., Halperin, E., Erlich, Y.: Identifying personal genomes by surname inference. Science 339(6117), 321–324 (2013)
11. Henne, B., Szongott, C., Smith, M.: Snapme if you can: Privacy threats of other peoples' geo-tagged media and what we can do about it. In: WiSec 2013: Proc. of the 6th ACM Conf. on Security and Privacy in Wireless and Mobile Networks, pp. 95–106 (2013)
12. Hoh, B., Gruteser, M., Xiong, H., Alrabady, A.: Enhancing security and privacy in trac-monitoring systems. IEEE Pervasive Computing 5(4), 38–46 (2006)
13. Huang, L., Yamane, H., Matsuura, K., Sezaki, K.: Silent cascade: Enhancing location privacy without communication QoS degradation. In: Clark, J.A., Paige, R.F., Polack, F.A.C., Brooke, P.J. (eds.) SPC 2006. LNCS, vol. 3934, pp. 165–180. Springer, Heidelberg (2006)
14. Humbert, M., Ayday, E., Hubaux, J.P., Telenti, A.: Addressing the concerns of the lacks family: Quantification of kin genomic privacy. In: CCS 2013: Proc. of the 20th ACM Conf. on Computer and Communications Security, pp. 1141–1152 (2013)
15. Krumm, J.: Inference attacks on location tracks. In: LaMarca, A., Langheinrich, M., Truong, K.N. (eds.) Pervasive 2007. LNCS, vol. 4480, pp. 127–143. Springer, Heidelberg (2007)
16. Krumm, J.: A survey of computational location privacy. Personal Ubiquitous Computing 13(6), 391–399 (2009)
17. Mislove, A., Viswanath, B., Gummadi, K.P., Druschel, P.: You are who you know: Inferring user profiles in online social networks. In: WSDM 2010: Proc. of the 3rd ACM Int'l Conf. on Web Search and Data Mining, pp. 251–260 (2010)
18. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. In: S&P 2009: Proc. of the 30th IEEE Symp. on Security and Privacy, pp. 173–187 (2009)
19. Shokri, R., Theodorakopoulos, G., Danezis, G., Hubaux, J.-P., Le Boudec, J.-Y.: Quantifying location privacy: The case of sporadic location exposure. In: Fischer-Hübner, S., Hopper, N. (eds.) PETS 2011. LNCS, vol. 6794, pp. 57–76. Springer, Heidelberg (2011)
20. Shokri, R., Theodorakopoulos, G., Le Boudec, J.Y., Hubaux, J.P.: Quantifying location privacy. In: S&P 2011: Proc. of the 32nd IEEE Symp. on Security and Privacy, pp. 247–262 (2011)
21. Srivatsa, M., Hicks, M.: Deanonymizing mobility traces: Using social network as a side-channel. In: CCS 2012: Proc. of the 19th ACM Conf. on Computer and Communications Security, pp. 628–637 (2012)
22. Vicente, C., Freni, D., Bettini, C., Jensen, C.S.: Location-related privacy in geo-social networks. IEEE Internet Computing 15(3), 20–27 (2011)
23. Vratonjic, N., Huguenin, K., Bindschaedler, V., Hubaux, J.P.: How others compromise your location privacy: The case of shared public IPs at hotspots. In: De Cristofaro, E., Wright, M. (eds.) PETS 2013. LNCS, vol. 7981, pp. 123–142. Springer, Heidelberg (2013)
24. Zheng, Y., Liu, L., Wang, L., Xie, X.: Learning transportation mode from raw GPS data for geographic applications on the web. In: WWW 2008: Proc. of the 17th ACM Int'l Conf. on World Wide Web, pp. 247–256 (2008)