Location Privacy with Probabilistic Co-Location Profiles [work in progress]

Jean-Pierre Hubaux, Kévin Huguenin, Roberto Pasqua, Alexandra-Mihaela Olteanu

Location Privacy with Probabilistic Co-Location Profiles [work in progress]

Jean-Pierre Hubaux[1], Kévin Huguenin[2], Roberto Pasqua[2], Alexandra-Mihaela Olteanu[1]
[1]*School of Computer and Communication Sciences, EPFL, 1015 Lausanne, Switzerland*
[2]*LAAS-CNRS, Université de Toulouse, 31400 Toulouse, France.*

1. Introduction

Location-based services, which rely on the localization capabilities of modern mobile devices (e.g., GPS), have become mainstream over the last decade. While very convenient, such services raise serious privacy issues, which have been extensively studied by the research community; and protection mechanisms, typically based on obfuscation techniques, have been proposed.

More recently, the (negative) effect of co-location information, that is the information that, at a given time instant, two or more users are at the same location, has been demonstrated [1]. Such information are, in fact, widely available to online service providers as they can be, for instance, extracted or inferred from pictures (by using face detection), tags on online social networks, and IP addresses (for users behind NATs).

In this work, we focus on the case where co-location profiles, that is the probability that two specific users are co-located at a given time of the day (e.g., "Alexandra and Roberto are co-located at noon in 80% of week days" or "colleagues are usually co-located during work hours"), are available, instead of specific co-location information (e.g., "On May 25th 2016 at noon, Alexandra and Roberto are co-located") as considered in previous works. Such co-location profile can typically be built from the history of user locations.

This short paper reports on the results of our preliminary investigation on how co-location profiles can be used to improve localization attacks, thus degrading users' location privacy. More specifically, we propose a simple inference algorithm that exploits co-location profiles and study its performance by relying on a dataset of mobility and proximity traces, namely the RealityMining dataset [2]. Our preliminary experimental results show that co-location profiles substantially affect location privacy.

2. System Model and Formalization

We consider a system of N mobile users who move in a given area of interest and an adversary, typically a service provider, who tries to infer the location of the users. We model the mobility of the users at discrete time and discrete locations: We consider T time instants and, at each time instant, a user is located in *one* of the M regions which compose the area of interest. The adversary *sporadically* observes the locations of the users: At each time instant, the adversary observes to the location of a *subset* of the users. In addition, the adversary has access to the *location* and *co-location* profiles of each user, under the form of the two following probability distributions: (1) the probability that the user is at a given location at a given time instant, and (2) the probability that two specific users are co-located at a given time instant. These profiles are typically periodic and specified over a period of a day.

3. Inference

We propose a simple algorithm to infer locations of users at a given time instant, given their co-location probabilistic profiles of the target user u and the actual locations of all the other users at the considered time instant. The inference algorithm picks the location of user u as the location l that maximize the following sum:

$$\sum_{v \text{ located at } l} \mathbb{P}(u, v \text{ are co-located}).$$

We compare this inferred location with the real location of u an we measure its accuracy as the proportion of correctly inferred locations. As a baseline, we consider an inference algorithm that simply picks the most frequent location for user u at the considered time instant.

Example:

- 5 users, A, B, C, D, E, $u = A, v \in \{B, C, D, E\}$
- Locations of users B,C,D,E are x, y, y, x resp.
- Real location of u is y
- Most common location of u at time t is x

v	$\mathbb{P}(u, v \text{ are co-located})$	Location of v (L)
B	0.2	x
C	0.19	y
D	0.15	y
E	0.1	x

- Score of location x is 0.3, score of location y is 0.34.

- We select y, and we compare this location with the actual location of u and results are
  - Heuristic ✓
  - Baseline ✗

4. Experimental Evaluation

In this section we present the dataset we used in the experimental evaluation and we report on the first experimental results of our study.

4.1. Dataset

To evaluate the potential of co-location profiles for infering users' locations, we rely on the Reality Mining dataset [2]. The reality Mining experiment was conducted between September 2004 and June 2005 by a research team at the MIT Media Laboratory. They used smartphones with pre-installed applications that record proximity and location data of each user, with respect to regular Bluetooth scans and GSM cell tower IDs respectively. They released an anonymous public version of this dataset that contains more than 400,000 hours of recording about the device usage behaviors. We use the ∼2 millions proximity events and ∼10 millions reported cell tower IDs (*i.e.* location events) for 106 human subjects involved in the experiment.

4.1.1. Construction. We suppose that the location of a user at time t is the reported cell tower ID. More precisely, for each hour of the day, for each user, we set her actual location as the most reported tower ID during the considered hour. This implies that only users who use the same mobile operator reported the same set of cell tower ID and, for this reason, we select the biggest subset of users with he same mobile operator, i.e. T-Mobile, eliminating users with partial reported information. Finally we retain 42 users for our study. We analyze the start and stop dates for each user to select the time interval of interest. Figure 1 shows the participation interval (e.g., the time interval between the first and the last event reported by a user) of each user.

We fix our interval I as the intersection of all participation intervals, i.e. 330 days, and we sample this interval with $\Delta t = 1$ hour to create a discrete successive time instants $t \in \{1, \ldots, T\}$ where we define the co-location events associated with reported locations.

4.1.2. Probabilistic co-location profiles. A co-location probabilistic profile is a pattern learned from the behavior of users or originated as a result of an analysis on background information about the relationships between users. More formally, a co-location probabilistic profile summarizes the probability of co-location at time t for a pair of users (u, v), so $\mathbb{P}(u, v \text{ are co-located} \mid u \leftrightarrow_{r_k} v)$. In order to generate the co-location probabilistic profiles we use the proximity data from Bluetooth scans.

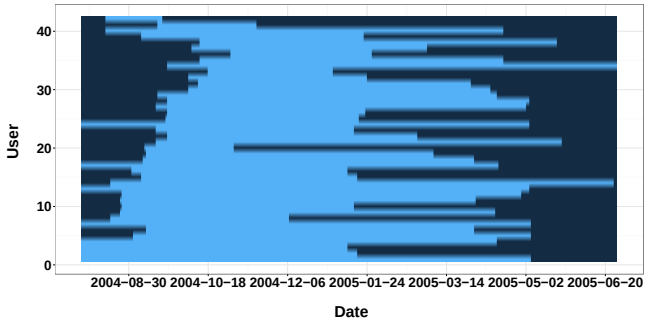

Figure 1. Start and stop dates, *light blue* lines identify the participation intervals. For the 42 users selected we represent the start date and the end date in a line.

Two users are co-located at a given discrete time instant t if and only if at least one of them reported at least a proximity event with the other user in this interval. Otherwise they are not co-located. Schematically it is a binary vector that has length T. We compute the probability of co-location for each pair of users at each hour of the day as the ratio of total co-location events in the same hour to numbers of active hour in T. A user is active in a hour $t \in T$ if and only if he reported at least a proximity event in this interval.

To summarize, Figure 3 shows the aggregation (arithmetic mean) of all probabilistic co-location profiles during each hour of the day.

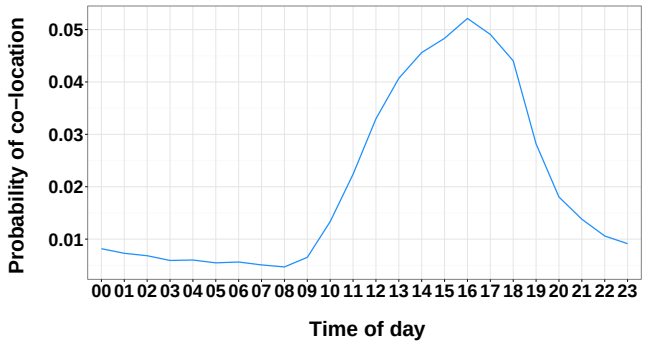

Figure 2. Probability of co-location averaged over all pairs of selected users in Reality Mining dataset.

To design our heuristic algorithm we use the co-location probabilistic profile of each pair of user.

4.2. Experimental Results

We run two algorithms for the 42 users and for every date and time ($\Delta t = 1$ hour) and we report the result in Figure 4.

We highlight that during the night it is difficult to improve the base line algorithm because it is more common for a user to have the same location and the

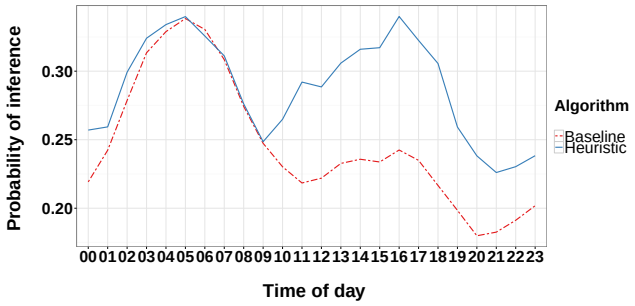

Figure 3. Result of heuristic algorithm compared to base line algorithm.

probability of co-location is very low. At the contrary, during the daytime we improve the base line result until a maximum of 40%. This result proves that implicit co-location information have a strong impact on the location privacy of a set of users.

5. Conclusion

In this report we present a simple co-location based inference attack. In particular, we mined the proximity and location information of users, respectively to produce implicit co-location probabilistic patterns and to extract the real location patterns of users. We demonstrated, by means of an inference location attack, the effect of these pieces of information on location privacy.

This is a preliminary result of a work in progress and we believe that we could improve our result by: (1) defining a more advanced inference algorithm which could take into account less location information, (2) implement an inference model based on Bayesian network, (3) use only probabilistic co-location profiles based on relationships between users (colleagues, friends, etc.) and (4) take into account also the mobility profiles of users, i.e. transition probabilities between regions.

References

[1] Alexandra-Mihaela Olteanu, Kévin Huguenin, Reza Shokri, Mathias Humbert, and Jean-Pierre Hubaux. Quantifying Interdependent Privacy Risks with Location Data. *IEEE Trans. on Mobile Computing (TMC)*, page 14, 2016. to appear.

[2] Nathan Eagle, Alex (Sandy) Pentland, and David Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences*, 106(36):15274–15278, 2009.

[3] Alexandra Mihaela Olteanu, Kévin Huguenin, Reza Shokri, and Jean-Pierre Hubaux. Quantifying the Effect of Co-location Information on Location Privacy. In *Proceedings of the 14th Privacy Enhancing Technologies Symposium*, Lecture Notes in Computer Science, pages 184–203, Berlin, 2014. Springer-Verlag Berlin.