Adaptive Personalized Privacy in Participatory Sensing

[Extended Abstract]

Berker Agir, Thanasis G. Papaioannou, Rammohan Narendula,
Karl Aberer and Jean-Pierre Hubaux
School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
Email: firstname.lastname@epfl.ch

ABSTRACT

The participatory sensing paradigm has interesting applications, e.g. electrosmog/air-pollution monitoring, carbon footprint estimation, etc., but raises serious privacy concerns. Existing static privacy-enabling approaches offer no privacy guarantees, while individual privacy requirements cannot be met. In this work, we propose an adaptive privacy protection scheme, in order to meet personalized location-privacy protection requirements and experimentally prove its effectiveness against static privacy-protection schemes.

Categories and Subject Descriptors

K.4.1 [**Public Policy Issues**]: Privacy; C.2.0 [**Computer-Communication Networks**]: General—*security and protection*

General Terms

Algorithms, Security

Keywords

Location Privacy, Utility, Data Hiding, Location Obfuscation

1. INTRODUCTION

The participatory sensing paradigm enables the vast collection of sensor data from privately-held sensory devices (e.g. mobile phones, vehicles, home appliances, etc.), and paves the ground for innovative applications of great social and business interest, such as monitoring air pollution and electrosmog, early earthquake detection, etc. However, users who are sensitive to their private information, such as their location (and inferred activities), are not expected to be willing to contribute to such systems. Thus, it is necessary to deploy some privacy-protection mechanisms. The usefulness of the sensed data for the application though is negatively affected by the privacy-protection mechanisms.

There exist location privacy-preserving mechanisms [2] employing techniques such as data hiding or location obfuscation. These techniques have been exploited by the research community for privacy-preserving participatory sensing before, but most of the existing approaches propose static protection frameworks, such as [1, 5]. Such approaches have two important problems: (*i*) They do not dynamically measure the privacy leakage based on the user actions. However, spatio-temporal correlation among user actions might reveal partial or full trajectories. (*ii*) They have a negative effect on the utility of the system, because the provided location privacy can be much higher in some cases than what a user actually wants, which can reduce data utility further.

In this work, we propose an innovative approach for continuous location-privacy protection in participatory sensing context, where the users estimate *locally* their expected location privacy and compare it with a personal privacy *threshold* before taking any privacy-protection action. Our adaptive scheme is light-weight, realistic and thus easily deployable on mobile devices. Based on real data traces, we experimentally show that this approach always satisfies the personal location-privacy protection requirements, when feasible.

2. ADAPTIVE PRIVACY SCHEME

In participatory sensing, mobile nodes (or just "nodes") sense their environment (e.g. temperature, speed, noise, etc.) and send their sensor data $< value, location, time >$ to a certain data-collection entity called Aggregation Server (AS). In our work, we consider the application of electrosmog monitoring by means of participatory sensing. We assume that the monitored area is partitioned into cells, while the time is slotted. The nodes are assumed to be honest, whereas the AS is semi-honest, meaning that it follows the protocols, does not collude with other entities and does not tamper with the system to obtain private information about the nodes. Additionally, the only background information the AS has on node mobility is the maximum possible speed.

Our approach to the personalized privacy-protection in this context is as follows: Nodes should be able to decide on privacy level requirement and configure privacy-protection mechanisms accordingly in an *adaptive* manner. This not only provides a better protection for nodes, but also decreases data utility loss. In light of this, it is important to enable nodes to locally assess their privacy leakage in real time: we use the *expected distortion* (ED) metric, proposed by Shokri *et al.* [4], which captures the uncertainty and the probability of success of the adversarial attack. ED of location inference for a node u at time t is given by:

$$ED(u,t) = \sum_{ol_t} D(\mathrm{loc}(u,t), ol_t) \cdot Pr(u, ol_t) \qquad (1)$$

where ol_t represents the observed locations of node u at time t. D is a normalized distance function in $[0,1]$, $\mathrm{loc}(u,t)$ gives the actual location of node u at time t. $Pr(u,ol_t)$ is the probability of ol_t being the actual location of node u. The result is in the interval $[0,1]$, where 0 means no privacy protection and 1 means full privacy protection.

We build a simple, yet effective location-privacy protection scheme based on the existing privacy-preserving techniques of location obfuscation and hiding. This light-weight scheme, depicted in Figure 1, works as follows: When a node has data to submit, it calls the location obfuscation module with the lowest obfuscation level, i.e. 0. Then, it provides the output of this module, which is a set of locations, to the privacy level estimation module. The estimation is compared against the node's privacy threshold θ. If θ is met, then the node submits the data to the AS. Otherwise, it increases the obfuscation level λ and repeats the process. If the desired privacy level is not reached even if predefined maximum λ is reached, then the data is not submitted.

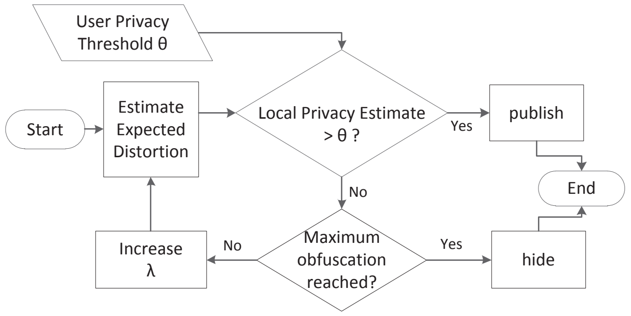

User Privacy Threshold θ
Start
Estimate Expected Distortion
Local Privacy Estimate > θ ?
Yes
publish
End
No
Maximum obfuscation reached?
Yes
hide
No
Increase λ

Figure 1: Adaptive Location-Privacy Protection Mechanism

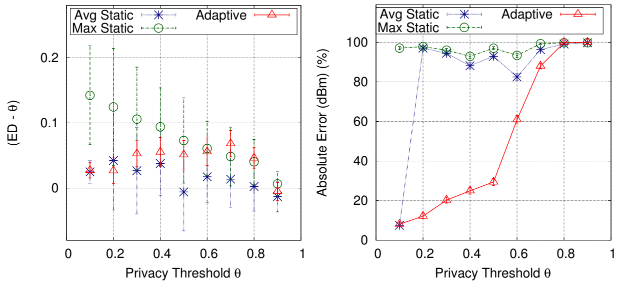
We calculate the expected privacy leakage locally by maintaining an event linkability graph at each node. Each vertex in this graph represents an event observable by the AS. An observable event corresponds to a data item associated to a location and a timestamp. The linkability graph is progressively constructed as new events are produced, by connecting events that are consecutive in time. Infeasible events are removed based on the geographical topology and the maximum possible speed. Each event in this graph is assigned a probability of being the node's actual location.

Based on the linkability graph, we use the Bayes rule to calculate the probability that an observed event corresponds to the actual location of a node. When a node wants to send data for the first time, a uniform probability $1/k$ is assigned to each corresponding vertex, where k is the number of vertices. As new vertices are added at a subsequent time instance, their probabilities to be genuine are calculated according to the Bayes rule. In case of vertex elimination due to infeasibility, the probabilities of all the affected vertices (i.e. parents or siblings) are updated and the update is propagated if necessary. After having calculated the probability of each observable event to be genuine up to the current time, a node calculates its expected distortion (ED) according to Equation (1). The complexity of this process is $O(M^2)$, where M is the number of vertices at a time step.

3. EVALUATION

We evaluate and compare our scheme to static policies experimentally, using 40 real traces collected during the Lausanne Data Collection Campaign [3] run by Nokia Research Center (Lausanne) and involving around 200 participants. We developed the experimental environment in C++ and the privacy evaluation is done using the Location Privacy Meter (LPM) by Shokri *et al.* [4], which is a comprehensive tool using various metrics including ED.

A static privacy-protection policy dictates a fixed hiding probability and a fixed obfuscation level for a node throughout its data emissions. To this end, we define two different static policies: (i) *Avg Static* is the policy which meets θ on the average, and (ii) *Max Static* meets θ most of the time.



(a) (b)

Figure 2: (a) Level of privacy achieved by adaptive vs. static policies for different privacy thresholds. (b) Percentaged absolute error (dBm) introduced by the various privacy policies.

As shown in Figure 2(a), the "Avg Static" policy violates θ almost half of the time (with 95% confidence interval), while the adaptive privacy-protection strategy almost always meets the privacy threshold requirement. Note that meeting $\theta = 0.9$ is very strict and sometimes infeasible with the employed location privacy-protection mechanisms. Also, observe that the adaptive strategy meets the various privacy thresholds more narrowly, as compared to the "Max Static" policy. As a result, the adaptive strategy is expected to deteriorate the utility of the participatory sensing application less than any static one, while satisfying the users' privacy requirements. Indeed, the absolute error as percentage of the data range introduced by the adaptive strategy is lower than the respective ones of the two static policies as shown in Figure 2(b). Also, Figure 2(b) clearly demonstrates the *trade-off* between utility and privacy. In terms of utility, our adaptive privacy protection strategy dominates any static strategies employing the same location-privacy protection techniques with us for any θ that render the participatory sensing application feasible.

4. CONCLUSION

We have defined a simple, yet effective, adaptive location-privacy protection scheme for participatory sensing and showed its effectiveness against static schemes based on real data traces. Our approach improves achievable utility, while satisfying the individual privacy requirements of users. As future work, we plan to consider the existence of application-related background information at the AS.

5. REFERENCES

[1] B. Gedik and L. Liu. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms. *Mobile Computing, IEEE Transactions on*, 7(1):1 –18, 2008.

[2] J. Krumm. A Survey of Computational Location Privacy. *Personal Ubiquitous Computing*, 13, 2009.

[3] Nokia Research Center. Lausanne Data Collection Campaign. `http://research.nokia.com/page/11367`.

[4] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux. Quantifying Location Privacy. In *Proc. of IEEE Symposium on Security and Privacy (S&P)*, 2011.

[5] X. Xiao and Y. Tao. Personalized Privacy Preservation. In *ACM SIGMOD*, 2006.