A Privacy Aware Approach for Participatory Sensing Systems

Dimitrios Tsolovos, Nicolas Anciaux, Valérie Issarny

A Privacy Aware Approach for Participatory Sensing Systems

Dimitrios Tsolovos
Inria & U. Versailles St-Q. en Yvelines
Versailles, France
dimitrios.tsolovos@inria.fr

Nicolas Anciaux
Inria & U. Versailles St-Q. en Yvelines
Versailles, France
nicolas.anciaux@inria.fr

Valérie Issarny
Inria
Paris, France
valerie.issarny@inria.fr

1 INTRODUCTION

Mobile Participatory Sensing (MPS) systems can be used to provide useful data gathered from mobile devices that would otherwise require expensive deployments of sensor networks which would also need to be maintained, thus, further adding to the overall cost. Data collected by these applications typically include the location and time as well as the actual measurement itself which can vary from environmental and health data, to any other sensor data that can be collected by the majority of modern smart phones. This data can be used purely for the personal benefit of the individual user (e.g. *"self quantification"*, systems) or they can be used to benefit every user of the application (e.g. traffic tracking).

This type of data is sensitive since it can be used to identify participants and infer important information such as their interests, location or possible medical conditions. Omitting directly identifiable data, such as the name of a user, from a data set is not enough to protect their privacy [12]. An adversary can extract user behavior patterns which, when combined with external knowledge can re-identify the users with high accuracy. MPS systems are by nature distributed and yet in most applications, participants are supposed to report the data they have collected to a central server and thus "re-centralizing" it. This approach assumes, by construction, that individuals do not question the honesty of the hosting company (and of its employees) nor its capacity to defeat severe attacks, since centralization creates in essence a massive honeypot.

Our objective is to provide a novel architecture for MPS systems which protects its users' privacy without sacrificing the utility of the system. The architecture should be designed such that participants can retain complete control over their data throughout the lifetime of the system and ensure that, an adversary performing a successful attack, will not be able to get access to all user data.

In the following sections, we present current privacy aware approaches for MPS systems. We discuss their limitations and derive a problem statement. Based on this problem statement, we extract the objectives that a privacy aware distributed MPS system should achieve and we provide an initial architecture in that respect.

2 PRIVACY IN TRADITIONAL MPS SYSTEMS

MPS systems are generally comprised of the following sub-systems: **a)** user registration, **b)** tasking, which includes the process of defining a task and finding suitable users to participate (task assignment) and collect relevant data, **c)** sensing and processing of local data, **d)** reporting, during which participants will transfer their collected data back to the server, **e)** global data processing, **f)** long term storage, and finally, **g)** presentation and end-user queries.

Each sub-system has different actors and consequently different types of adversaries. Typical MPS systems, generally have four types of actors. The server administrators, the tasking entities, the participants and the end-users. During the tasking process, a tasker defines a task by describing the points of interest, the types of measurements, the temporal scope of the task and more. Moreover, they define a function that needs to be applied on the collected data. An adversarial tasking entity might create a task that targets specific people by providing a very restrictive task description. Once the task is distributed and the data collected by the users, they will be called to report it to the server. Generally, this is where Privacy Enhancing Techniques (PETs) such as k-anonymity, differential privacy and data aggregation techniques are applied. Data need to be anonymized before reaching the server. Once on the server, a corrupted administrator can use the data to expose the participants privacy. Once the final results are published, end-users might be given the ability to perform queries. Their own privacy can also be exposed simply by inferring their interests based on their queries.

Privacy threats in MPS systems can be grouped into three categories: **a)** *Data snooping*, where adversaries might attempt to gain access to sensitive participant data by targeting specific participants. **b)** *Data inference*, where an adversary might attempt to extract additional information from the participants by based on the collected data and external information. **c)** *integrity of the system*, where adversarial actors might attempt to corrupt the system.

In recent studies, a trusted entity in the architecture, will be responsible for anonymizing user data using some PET. In [2] and [8] that is an anonymization server while in [6] that role is held by the mobile phone service provider. In [3], participants are called to exchange their measurements with each other before reporting them back to the server. This way, the link between the participant and the data is broken. This alone is not sufficient and it needs to be coupled with other PETs as this link can be reestablished by analyzing a participant's history and combining it with external knowledge. For a survey of relevant studies see [1].

Current studies assume that central points of the system are trusted with user data. This in many cases is not a realistic scenario since companies like Google collect massive amounts of user data while their incentives lie in giving advertisers more relevant views to their ads and user privacy becomes secondary. Additionally, they usually employ techniques such as k-anonymity or differential privacy, which reduce data utility. In general, centralized approaches fail to provide users with guarantees that their data will not be misused either intentionally, by negligence or in case of a successful attack against the server.

3 OUR APPROACH

A privacy aware architecture for MPS systems should enforce that the benefit-to-cost ratio of a successful attack is reduced and additionally, the cost of attacking multiple user data should increase (at least) linearly with the number of users. It should enable users to have complete control over their own data. The user should be the

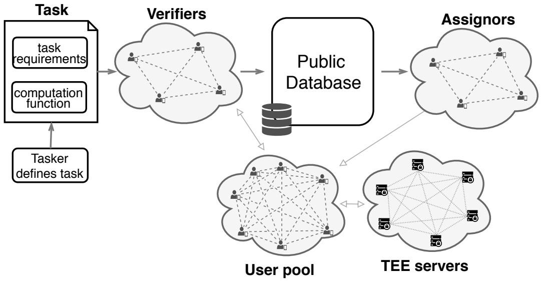

Figure 1: A high level overview of our proposal.

only one who has her own data in the clear and should have control over its deletion, usages and more. Finally, the architecture should not sacrifice main functionalities (such as optimal task assignment, rich computations, etc.) to achieve privacy.

Our approach relies on users keeping collected data on their own devices. This prevents an adversary from getting access to all user data with a single attack. This does not ensure that the benefit-to-cost ratio is reduced. An exploit that can “break” a device can be distributed to multiple devices without increasing the cost as seen in the recent Spectre and Meltdown attacks. The threats remain the same as in the centralized approach but their scope is different. The move to a decentralized setting ensures that users keep control of their data and can choose how they will be used. Additionally, on our approach we utilize devices equipped with a Trusted Execution Environment (TEE) [7, 9]. A TEE is a secure area of the main processor that guarantees code and data loaded in it will be protected with respect to confidentiality and integrity.

The main challenge when moving to a decentralized approach is achieving computations over global data. There are multiple techniques in the literature enabling privacy preserving distributed computations. These include Secure Multiparty Computations (SMC) [11] which does not scale well with the number of parties and gossip protocols [5] which can not be used for general computations. With the use of TEE enabled devices we can achieve fast and secure computations. Our objective is to ensure that an adversary breaking one device, will not be able to perform the same attack to break the other devices in the system. This will lead to the benefit-to-cost ratio being linearly reduced with the number of devices. Moreover, the attestation property that some TEEs offer can ensure that the result is a product of the specified function and which was not tampered with. In [4] the authors propose an architecture where users hold devices equipped with secure hardware with the goal of protecting the location privacy of the participants. Moreover, they propose protocols for efficient data aggregation. However, optimal task assignment is not considered in this study and location is not the only privacy sensitive information that participants provide in such systems.

In Figure 1 we present a high level architecture of our proposal. The key features of our architecture are the following: **a)** A user defines a task in the form of a smart contract and publishes it in an immutable, append and read only public database. **b)** A selection mechanism will select a subset of the available users (verifiers) who will test the task for well known tasking threats. **c)** If it passes the verification process, it can be distributed to the users. **d)** Another selection process selects a subset of users (assignors) who will assign the task to the available users based on their compatibility with the task description. **e)** The users will then perform the required measurements according to the task and save the collected data on their devices. **f)** With the assistance of a set of TEE enabled devices, the participants will collaborate to execute the defined function.

In this initial architecture user data is held on the user devices. When participants need to share their collected data in order to perform computations, the use of TEEs can help in keeping clear data away from adversaries. The two selection mechanisms we introduce are going to assume the role that a central server would have in a typical MPS system. Participants will need to collaborate in order to ensure that task verification and distribution is optimal while ensuring that user data is not collected on a single location. Ensuring the integrity and confidentiality of these two mechanisms is not a trivial task, even with the use of TEEs, since these can suffer from side channel attacks which can lead on massive data leaks, thus impeding confidentiality, while the integrity of the result is not guaranteed as well when computations are distributed among multiple devices.

4 FUTURE WORK

In this paper, we argued that, since MPS systems are inherently distributed, keeping them distributed can bring benefits to user privacy. Contrarily, it also brings a lot of challenges that will need to be overcome. Moving forward, we need to refine this initial architecture and prove that it can achieve our specific goals. Moreover we need to address data oriented tasks on TEEs. In particular, optimal task assignment and global computations over time series. An interesting challenge would also be to formulate the tasks as smart contracts, which, along with TEE attestation techniques, can provide integrity to the system. Finally, we plan to build part of this architecture and evaluate it over real user data collected by the Ambiciti project [10].

ACKNOWLEDGEMENTS

This work is part of a Ph.D. thesis funded by the Inria project City-Lab@Inria co-supervised by Nicolas Anciaux and Valérie Issarny.

REFERENCES

[1] Delphine Christin. 2016. Privacy in mobile participatory sensing: current trends and future challenges. *Journal of Systems and Software* 116 (2016), 57–68.
[2] Cory Cornelius et al. 2008. Anonysense: privacy-aware people-centric sensing. In *6th international conference on Mobile systems, applications, and services.* ACM.
[3] Delphine Christin et al. 2011. Privacy-preserving collaborative path hiding for participatory sensing applications. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on.* IEEE, 341–350.
[4] Dai Hai Ton That et al. 2016. PAMPAS: Privacy-Aware Mobile Participatory Sensing Using Secure Probes. In *ACM SSDBM 2016.*
[5] David Kempe et al. 2003. Gossip-based computation of aggregate information. In *IEEE Symposium on Foundations of Computer Science.*
[6] Hien To et al. 2014. A framework for protecting worker location privacy in spatial crowdsourcing. *VLDB Endowment* 7, 10 (2014).
[7] Ittai Anati et al. 2013. Innovative technology for CPU based attestation and sealing. In *2nd international workshop on hardware and architectural support for security and privacy*, Vol. 13. ACM New York, NY, USA.
[8] Khuong Vu et al. 2012. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *IEEE INFOCOM.* IEEE, 2399–2407.
[9] Tiago Alves et al. 2014. TrustZone: Integrated Hardware and Software Security-Enabling Trusted Computing in Embedded Systems (July 2004).
[10] Ventura Raphaël et al. 2017. Estimation of urban noise with the assimilation of observations crowdsensed by the mobile application Ambiciti. (2017).
[11] Oded Goldreich. 1998. Secure multi-party computation. *Manuscript.* (1998).
[12] John Krumm. 2007. Inference attacks on location tracks. In *International Conference on Pervasive Computing.* Springer, 127–143.