FOUGERE: User-Centric Location Privacy in Mobile Crowdsourcing Apps

Lakhdar Meftah, Romain Rouvoy, Isabelle Chrisment

FOUGERE: User-Centric Location Privacy in Mobile Crowdsourcing Apps

Lakhdar Meftah[1], Romain Rouvoy[2], and Isabelle Chrisment[3]

[1] Inria / Univ. Lille, France lakhdar.meftah@inria.fr
[2] Univ. Lille / IUF / Inria, France romain.rouvoy@inria.fr
[3] LORIA-TELECOM Nancy / Univ. Lorraine, France
isabelle.chrisment@loria.fr

Abstract. Mobile crowdsourcing is being increasingly used by industrial and research communities to build realistic datasets. By leveraging the capabilities of mobile devices, mobile crowdsourcing apps can be used to track participants' activity and to collect insightful reports from the environment (*e.g.*, air quality, network quality). However, most of existing crowdsourced datasets systematically tag data samples with time and location stamps, which may inevitably lead to user privacy leaks by discarding sensitive information.
This paper addresses this critical limitation of the state of the art by proposing a software library that improves user privacy without compromising the overall quality of the crowdsourced datasets. We propose a decentralized approach, named FOUGERE, to convey data samples from user devices to third-party servers. By introducing an *a priori* data anonymization process, we show that FOUGERE defeats state-of-the-art location-based privacy attacks with little impact on the quality of crowdsourced datasets.

Keywords: Location Privacy · Mobile Crowdsourcing · LPPM.

1 Introduction

Mobile crowdsourcing platforms and applications (or apps) are being widely used to collect datasets in the field for both industrial and research purposes [2,6,31]. By relying on a crowd of user devices, mobile crowdsourcing delivers an engaging solution to collect insightful reports from the wild. However, the design of such platforms presents some critical challenges related to the management of users, also known as *workers*. In particular, the privacy of the workers is often underestimated by the crowdsourcing platforms and it often fails to be addressed effectively in practice [25].

While data anonymization is commonly achieved *a posteriori* on the server side [7,16,20,22], this approach is subject to adversarial attacks, even when protocols for the communication and the data storage are claimed to be secured [11,12]. Furthermore, the workers may be reluctant to share *Sensitive Personal Information* (SPI) with third parties (*e.g.*, students contributing to

a crowdsourcing campaign initiated by a professor). Gaining the confidence of workers is extremely difficult and we argue in this paper that the adoption of *a priori* data anonymization mechanisms contributes to delivering a trustable component to better mitigate privacy leaks in the data shared by workers.

For example, the worker's location is not only the most requested but also the most sensitive data collected by mobile crowdsourcing platforms [3]. Our scheme therefore explores the physical proximity of workers to agree on a dissemination strategy for reporting the crowdsourced data. By altering the link between workers and data *consumers* on the server, our approach intends to mix data contributed by several workers within a collaborative data flow that exhibit similar crowd-scale properties and without discarding any SPI. In particular, we propose a system-level service that acts as a proxy within the mobile device for sharing crowdsourced data and from which workers can control their privacy settings. FOUGERE is our implementation of this anonymization scheme and is available as an open source library[4] that can be used by legacy mobile crowdsourcing apps. We illustrate the benefits of FOUGERE by integrating it within the state-of-the-art MOBIPERF mobile crowdsourcing app as well as the APISENSE mobile crowdsourcing platform. We evaluate the effectiveness and the impact of our anonymization scheme on these two mobile crowdsourcing systems by deploying and orchestrating a crowd of 15 emulated mobile devices. More precisely, we replay the SFCABS cab mobility traces [30] and we show that FOUGERE defeats state-of-the-art privacy attacks [18,24,26] with little impact on the quality of the resulting datasets.

The remainder of this paper is organized as follows. Section 2 gives a background on mobile crowdsourcing platforms and discusses the related work in the areas of mobile crowdsourcing and location-based privacy. Section 3 provides an overview of the privacy threats in crowdsourcing apps and platforms. Section 4 introduces our anonymization scheme and the integration of LPPMs to increase the workers' privacy. Section 5 describes the implementation of the FOUGERE open source library on Android. Section 6 introduces our evaluation protocol of FOUGERE on the MOBIPERF mobile crowdsourcing app and discusses the results we obtained on an experimental setup involving 15 emulated workers. Section 7 discusses the threats to validity of our contribution. Finally, Section 8 concludes on this paper.

2 Related Work

Thanks to the wide adoption of mobile devices, mobile crowdsourcing has emerged as a convenient approach to gather meaningful and scalable environmental datasets by involving citizens in the process of performing measurements in the wild [2,6,22,31]. While the development of mobile crowdsourcing apps is clearly leveraged by the *Software Development Kits* (SDK) made available by Android and iOS, mobile crowdsourcing platforms are bringing another level of abstraction to ease the design and the deployment of mobile crowdsourcing campaigns [4,8,13,20].

[4] `https://github.com/m3ftah/fougere`

As depicted in Figure 1, mobile crowdsourcing campaigns typically consist of several stages: *i)* the description of the data to be crowdsourced, *ii)* the deployment and the gathering of the dataset in the wild, *iii)* the aggregation and storage of datasets in the Cloud, *iv)* the processing and *v)* publication of the campaign results. However, along all these stages, SPI can be conveyed by the platform and potentially be subject to attacks from adversaries, therefore motivating the development of a better privacy support.

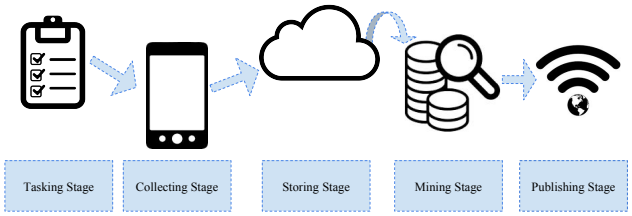

Fig. 1. Anatomy of a mobile crowdsourcing campaign

Location privacy protection mechanisms (LPPMs) are particularly interesting to limit user privacy leaks [5]. A large body of the related work has been devoted towards the latest stages of mobile crowdsourcing campaigns by improving the privacy properties of datasets once uploaded to remote servers [23,28,35]. These techniques contribute to preserving the privacy of workers while limiting the impact on the quality of the resulting dataset. However, raw datasets stored on a remote server may be leaked through security breaches.

Collaborative privacy-preserving location-based services. In the domain of *Location-Based Services* (LBS), some privacy protection mechanisms can be adapted to mobile crowdsourcing platforms. In particular, we consider the solutions where users collaborate to hide information from the server [10,29,32,33], which share similarities with FOUGERE. In particular, Show *et al.* [10] use communication over *peer-to-peer* (P2P) protocols, Shokri *et al.* [32] use WiFi *Access Point* connection, and finally, Shokri *et al.* [33] and Peng *et al.* [29] propose to use Wi-Fi Direct communications. Yet, such approaches are not widely adopted by LBS solutions as they fail to demonstrate their effectiveness in a realistic deployment.

Privacy in mobile crowdsourcing platforms. Mobile crowdsourcing platforms are actively working on privacy protection mechanisms [3,19]. In particular, Cornelius *et al.* [13] have proposed ANONYSENSE: a mobile platform for opportunistic sensing. Because the server hosting the collected dataset can trace the worker's wireless access points, they propose to use an anonymization network to hide worker locations, they rely on a third-party server for routing the data. ANONYSENSE also supports reporting data with a statistical guarantee

of k-anonymity. The workers' data are blurred and combined before being reported to the remote server. While their approach hides workers from the server, it exposes them to a third-party server that has to be trusted by the workers. Thus, introducing a single point of failure. Das *et al.* [14] present PRISM: a platform for remote sensing. They use a sandbox to prevent mobile apps from using mobile sensors. Adversaries can still collect geotagged data from workers and apply privacy de-anonymization attacks on the dataset. As discussed in [14], both ANONYSENSE and PRISM suffer from similar privacy leaks as the mobile app collects data local sensors made available by their mobile device, allowing data to be linked to the worker identifier. Hu *et al.* [21] present a collaborative privacy-preserving platform called HP3, which uses social networks to hide workers from the server. In their approach, they rely on third-party servers (the social network) that can store all the exchanged locations along with workers identifiers.

Synthesis. To the best of our knowledge, the state of the art fails to appropriately address the anonymization schemes along the earliest stages of a mobile crowdsourcing campaign in order to limit potential privacy threats. Therefore, in this paper, we intend to address this limitation by proposing an approach that leverages existing privacy protection mechanisms from the mobile device by providing the first decentralized dissemination to adjust location privacy in mobile crowdsourcing systems.

3 Privacy Threats in Mobile Crowdsourcing Systems

This section discusses the potential threats in mobile crowdsourcing systems along 3 axes: the *system model*, the *sensitive personal information*, and the *known location-based attacks*.

Mobile crowdsourcing system model. The architecture we consider is a mobile crowdsourcing campaign that involves three components, namely, *mobile devices*, *crowdsourcing apps*, and *storage servers*.

We consider that the mobile *crowdsourcing apps* can be trusted as we believe that the owner of the mobile crowdsourcing app or platform is interested in gathering insightful datasets with the consent of workers, especially if this mobile app is open sourced.

However, we consider that the *storage server* can be compromised and reveal some sensitive personal information on behalf of the owner and the workers. For example, no matter if they are deployed in the cloud or on-premise, the remote storage servers may suffer from security leaks that can be exploited by an adversary. Furthermore, storing the crowdsourced data on the server must comply with *The EU General Data Protection Regulation (GDPR)* and the *Privacy Act of 1974* of the USA. With crowdsourced data, it is difficult to comply with the regulations, for example: giving the users the right to delete their own data whenever they want. FOUGERE does resolve issues related to these regulations as it does not store personal identifiers on the server side.

Sensitive personal information in mobile crowdsourcing. The goal of a mobile crowdsourcing system is to gather a very large volume of data from measurements produced by third-party workers. These workers are recruited by the owner of the mobile crowdsourcing system to upload crowdsourced data through a dedicated mobile app or device. However, existing mobile crowdsourcing systems may collected some sensitive personal information.

In particular, we identify 4 categories of *sensitive personal information* (SPI) that might be exploited by attackers:

Identifiers group all persistent or transient identifiers that can take the form of a device ID (IMEI) or Google account ID, for example, to explicitly identify a worker from the perspective of a mobile crowdsourcing system. However, such identifiers may directly name the worker or be used to perform context linking attacks by combining several measurements;

Point of Interests (POI) gather all the forms of geolocated data that can deliver some spatial information on the location of a worker. This includes GPS locations, but also places check-in, cell tower ID or location, which are used by some systems to produce maps from crowdsourced measurements. However, these POI may also reveal the home, office, shopping and/or leisure locations of workers that can uniquely identify them [17];

Routines concern any information that can be use to capture a recurrent activity of a worker. This category of SPI covers in particular any form of timestamp, no matter the format and the precision. While this precious information often appears as harmless, it may also be used by context linking attacks to group crowdsourced data and observe correlation along time (*e.g.*, nights, week-ends);

Markers finally focus on information whose entropy in terms of values can be exploited to detect outlier workers and thus be indirectly used as an identifier by an attacker. There can be a wide diversity of such markers depending on the purpose of the mobile crowdsourcing system. For example, in the case of MobiPerf, the properties of device manufacturer, model, OS version and network carrier can be considered as unique if a worker uses some original/old mobile device.

Location privacy attacks. Similarly to [37], we consider that the adversary can exploit two dimensions of knowledge: *temporal information* and *context information*.

In the context of mobile crowdsourcing systems, *temporal information* refers to the capability of the adversary to access a history of crowdsourced data—*i.e.*, several measurements reported by a single worker. In the case of a compromised storage server (or connection to the storage server), such assumption holds as the attacker gets access to sufficiently large volume of crowdsourced data to build some temporal knowledge.

Beyond spatio-temporal information, *context information* refers to any additional information that an attacker can exploit. This covers embedded knowledge that is included in the crowdsourced dataset (*e.g.*, markers) or side knowledge

that an attacker can obtain from other information sources (*e.g.*, the number of involved workers).

4 FOURGERE: Empowering Workers with LPPMs

To overcome the above privacy threats and strengthen the location privacy of workers, this paper introduces a new middleware library, named FOUGERE, which acts as an embedded proxy to anonymize and disseminate the workers' crowdsourced data across the network. This section introduces the key design principles we adopted, a description of how crowdsourced data flows across multiple devices, as well as the core *Location Privacy Protection Mechanisms* (LPPMs) that are provided by FOUGERE.

Collaborating with apps & workers. In order to be trusted and gather a large crowd of workers, we assume that mobile crowdsourcing apps and platforms are doing their best to enforce privacy and security support. However, developers are not necessarily aware of privacy threats and implementing a comprehensive support for such a support might be time-consuming and error-prone. FOUGERE therefore offers mobile crowdsourcing apps the possibility to offload the management of the worker privacy settings and the data dissemination across the network, thus letting developers focus on the core business of the mobile app. More specifically, FOUGERE offers the workers control over worker's privacy preferences, thus providing a preference panel to *i)* explore the list of mobile crowdsourcing apps and respective SPI, *ii)* monitor and control the volume of crowdsourced data reported by each app, and *iii)* configure the list of LPPMs to be enforced by a given mobile crowdsourcing app.

By following these principles, FOUGERE can collaborate with the mobile app and the worker to ensure the anonymization and the dissemination of crowdsourced data. Figure 2 overviews these principles and illustrates how a mobile app can disseminate crowdsourced data without and with FOUGERE. In particular, mobile crowdsourcing apps that do not fulfill the design principles—or do not integrate FOUGERE—will upload crowdsourced data directly to the remote server, thus exposing the workers to the privacy threats introduced in Section 3. By integrating FOUGERE, any mobile crowdsourcing app simply delegates the data dissemination to the library. FOUGERE enforces the worker's privacy settings and applies the appropriate LPPMs to the forwarded data. Such mechanisms include *privacy filters* (to discard the data), *privacy distortions* (to alter the data) and *privacy aggregation* (to group the data).

Enabling crowdsourced dissemination. If the crowdsourced data has not been discarded by one of the configured LPPMs, FOUGERE stores a message for dissemination that is composed of *i)* a *payload*, *ii)* a *configuration* of remote LPPMs, *iii)* a *bloom filter* of forwarder devices, and *iv)* a *time-to-live* (TTL) for the dissemination process. While the *payload* refers to the crowdsourced data, which has eventually been altered by the local LPPMs, the message also includes some *configuration* parameters for LPPMs that can be executed by

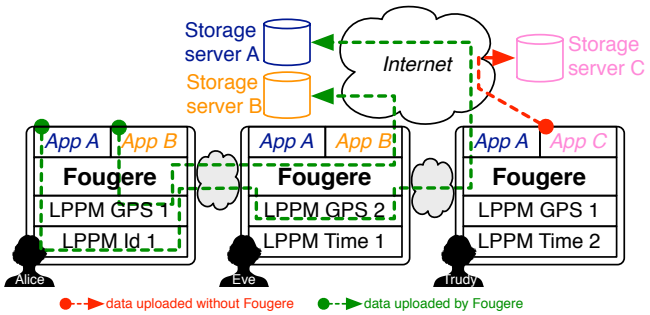

Fig. 2. Overview of FOUGERE

remote instances of FOUGERE (*e.g.*, replacing the location of the source by the location of the forwarder). In order to avoid a given message to be forwarded by the same set of mobile devices, FOUGERE also includes a bloom filter that encodes the list of forwarder nodes, without discarding their identifier. The *bloom filter* is configured with a false positive probability of 0.1 and a number of expected elements equals to the TTL. Finally, the message encloses a *TTL* to define the numbers of workers hops requested by the worker to disseminate the message.

FOUGERE filters out the known workers by querying the bloom filter, and randomly picks and forwards the message to one the remaining nodes. Upon receiving such a message, a remote FOUGERE node eventually applies the LPPM listed in the configuration before checking the TTL. If the TTL equals 0, then FOUGERE stores the payload for being forwarded by the mobile crowdsourcing app to the remote storage server. Otherwise, FOUGERE decreases the TTL, adds its own identifier to the bloom filter, and stores the resulting message for further dissemination.

Mobile crowdsourcing apps share similarities with *Delay Tolerant Networks* (DTN) by considering that the crowdsourced data does not have to be immediately uploaded to the remote server and can tolerate delays ranging from minutes to hours. We exploit this property to adopt a multi-hop forwarding scheme in FOUGERE, which ensures that at least k neighboring devices with the same mobile app are also potentially collecting data in the same area, thus preventing the worker to be spotted as an outlier.

Furthermore, FOUGERE complements existing privacy-preserving mechanisms, like the TOR anonymity network, which can also be used by FOUGERE to upload the crowdsourced data to the remote server. Using TOR, therefore, hides workers from the remote server, but it loses the physical proximity information that is useful for local LPPMs. For example, when an isolated worker is contributing from within the countryside, she can still report data using TOR but she will remain exposed to location privacy attacks.

Controlling LPPMs from devices. In order to give the worker more control over her own data, FOUGERE includes several LPPMs that can be configured by

the worker to decide upon the quality and the volume of crowdsourced data to be obfuscated. In particular, we consider 3 classes of LPPMs: *filters*, *distortions*, and *aggregations*, which can be implemented within a mobile device and used to obfuscate one of the SPI of the user.

Privacy Filters are a group of LPPMs that can decide autonomously if a crowdsourced data can be shared with the crowdsourcing platform or not. For example, a LocationFilter applies to *points of interests* and can be configured by the worker to define *white areas* or *black areas* that delimit zones where the mobile crowdsourcing app can or cannot collect data, respectively. Similarly, a TimeFilter rather applies on *routines* and is used with configured periods along which a mobile crowdsourcing app can or cannot collect data. Finally, a QuotaFilter is a more generic filter that can accept a worker-defined quota of crowdsourced data to be uploaded before discarding once this quota is reached.

Privacy Distortions are another class of LPPMs that can modify the value of an enclosed SPI in the crowdsourced data to be shared. For example, a IdentifierDistortion will change the value of an identifier at a given frequency (every request, hour, day), while a location distortion adds a controlled random noise to the worker's location (depending on radius r with a level of privacy that depends on r) into the reported coordinates [15].

Privacy Aggregations reflect the last class of LPPMs that are supported by FOUGERE and propose to delay the dissemination of crowdsourced data by grouping them along a given criteria. For example, a TimeAggregation will group data per hour and apply an aggregation operator (like the average, the median, the min or the max) to the enclosed timestamp in order to report the same value for all the aggregated samples before reporting them. A MarkerAggregation is an example of remote LPPMs that will be encapsulated with the crowdsourced data and wait for a given marker (*e.g.*, the ISP name) to appear at least k times before being uploaded. This LPPM is an example of a distributed implementation of the k-anonymity algorithm [9,34] that we can apply on a wide diversity of SPI, including GPS coordinates.

Summary. By combining an opportunistic dissemination scheme with worker-defined LPPMs, FOUGERE aims at leveraging the privacy properties of legacy mobile crowdsourcing apps and platforms. Before assessing the efficiency of FOUGERE, we now report on the implementation of these principles on the Android platform.

5 Implementation Details on Android

On Android, FOUGERE is packaged as an open source library that deploys system service within the mobile device of a worker. This system service currently builds on the Wi-Fi Direct network interface to exchange crowdsourced data between nearby devices of workers. It can be shared by multiple crowdsourcing apps of a given device to centralize the control of privacy settings, which are exposed to the worker as a dedicated preference panel. Thanks to its modular architecture,

FOUGERE can be further extended with additional LPPMs, which are not covered by this paper.

Application programming interface. Any mobile crowdsourcing app can integrate FOUGERE through a simple API that exposes the following operations:
`hasFields(...)` is called by the mobile crowdsourcing app to declare any SPI as a `PrivacyField`, that refers the classes `IDENTIFIER`, `POI`, `ROUTINE`, and `MARKER`;
`forward(...)` enlists a task in charge of uploading a crowdsourced data sample to the remote server when the TTL expires;
`send(...)` delegates the dissemination of a crowdsourced data to FOUGERE.

Opportunistic dissemination. The current implementation of the FOUGERE dissemination module builds on the WiFi-Direct technology to discover nearby devices. When a mobile crowdsourcing app forwards a message, FOUGERE triggers the configured LPPMs and accumulates the data in the forwarding queue. For each data accumulated in the forwarding queue, FOUGERE picks a random peer that has never received this data and forwards it.

If the message reaches the configured number of device hops ($ttl = 0$), then the forwarded data is placed in an uploading queue, which will be emptied as soon as the remote mobile crowdsourcing app runs by invoking the upload handler registered by the app.

LPPM integration. FOUGERE combines the implementation of a decentralized dissemination scheme with the integration of LPPMs that can filter out data or alter its content depending on the worker's privacy settings. More generally, FOUGERE intends to leverage the integration of additional LPPMs to better control the data uploaded by any compatible crowdsourcing app. FOUGERE organizes these LPPMs along the 4 categories of SPI it supports. An LPPM complies to am interface `Lppm<T extends PrivacyField>` that declares the category `T` of SPI it considers and implements a method to apply a privacy mechanism on the uploaded data, which eventually returns the anonymized data to be further processed by FOUGERE.

In order to effectively apply the worker's privacy settings, FOUGERE operates by first applying the privacy filters, before proceeding with privacy distortions and finally privacy aggregations. In addition to that, privacy distortions and aggregations can also be triggered remotely to implement decentralized algorithms that build on neighboring samples to increase the privacy of workers [1].

6 Evaluations of FOUGERE

6.1 Evaluation Protocol

Beyond the challenges related to the integration in legacy mobile crowdsourcing systems, FOUGERE intends to deliver an efficient adoption of LPPMs in a decentralized context. The validation of such a capability requires consideration of a realistic deployment of mobile devices in order to assess the benefits of FOUGERE. Given that we are interested in providing a proof of feasibility for

FOUGERE, we are not interested in simulating the behaviour of LPPMs, but rather in assessing the reference implementation of FOUGERE. However, testing mobile applications that make use of opportunistic communications is hard to achieve and reproduce with real mobile devices. We propose to deploy a cluster of emulated devices to reproduce the behavior of a crowd of workers who contribute to a mobile crowdsourcing campaign. We use mobility datasets that are publicly available to control the emulated devices and we collect their interactions to trace their actions *a posteriori*. The crowdsourced dataset collected on the remote server are evaluated by the LPM^2 toolkit [32] to evaluate the preservation of workers' privacy. By adopting such an empirical validation, we can evaluate real applications integrating FOUGERE and we can observe the impact of changing the parameters of FOUGERE (number of hops, LPPMs' specific parameters).

In the remainder of this section, we select the legacy MOBIPERF [22] mobile app as the mobile crowdsourcing app that we considered to assess FOUGERE.

Emulating crowds of workers The assessment of our opportunistic dissemination scheme and the associated LPPMs requires consideration of a crowd of workers who installed a mobile crowdsourcing app that integrates FOUGERE. While running an emulator on a single machine is rather resource-consuming and cannot scale, we propose to consider the deployment of a cluster of servers to host multiple Android emulators. As Android emulators do not provide any support for ad hoc communications, such as WiFi-Direct, we use ANDROFLEET [27] to control the discovery of nearby devices within a cluster of emulators.

Controlling crowds of workers. To assess the efficiency of FOUGERE in the ANDROFLEET cluster, the emulated devices are required to be controlled in order to update their location and eventually internal state, to reproduce the mobility of a crowd of workers. While the choice of such a mobility dataset might be challenging depending on the category of mobile crowdsourcing app, we use the `epfl/mobility` dataset that is publicly available from CRAWDAD [30] to emulate 15 workers who are performing network measurements with the MOBIPERF mobile app. The crowdsourced dataset contains network measurements reported every 5 minutes by the workers moving in the San Francisco bay area.

Attacking crowdsourced datasets. To evaluate the impact of FOUGERE on the privacy of workers, we use the LPM^2 toolkit [32], which is a state-of-the-art tool for measuring location privacy. In particular, LPM^2 covers the evaluation of the LPPMs that are supported by FOUGERE, like the obfuscation mechanisms including perturbations (adding noise), reducing precision, location hiding. To validate FOUGERE against privacy attacks, for each configuration, we run an experiment that follows these steps:

1. Run ANDROFLEET with MOBIPERF and FOUGERE (incl. privacy settings),
2. Assign tasks to workers during 3 days, and wait 4 more days for the data dissemination to complete,
3. Gather the logs of data exchanges between workers to evaluate the opportunistic dissemination scheme,

4. Retrieve all the raw crowdsourced data stored on the remote server,
5. Construct the adversary knowledge by tagging the crowdsourced data of one worker (as required by LPM^2),
6. Evaluate the privacy support of FOUGERE with the LPM^2 toolkit,
7. Report on performance, utility, robustness and uncertainty, which are the parameters proposed by Verykios *et al.* [36] to assess LPPMs.

6.2 Empirical Evaluation

In this section, we instantiate the above experimental protocol to assess FOUGERE as a practical support to improve the location privacy of workers.

Experimental Setup In particular, thanks to the ANDROFLEET [27] emulation platform, we can reproduce the execution of a deployment of 15 mobile instances emulating a one-week crowdsourcing campaign, thus proposing a realistic input dataset to evaluate FOUGERE. Then, we compare the behaviors of 6 configurations of the MOBIPERF app:

1- VANILLA refers to the reference implementation of the MOBIPERF Android app, as it can be downloaded from *http://www.mobiperf.com*. This configuration is used to demonstrate the vulnerability of legacy mobile crowdsourcing apps with regards to potential privacy threats. It is also used as a witness to evaluate the benefits of the other configurations including FOUGERE;

2- FOUGERE *with no LPPM* refers to the extension of MOBIPERF with the FOUGERE library. This configuration is used to isolate the properties of our opportunistic dissemination schemes independently of the impact of LPPMs. In particular, we consider the following worker configurations for the number of required hops to disseminate the crowdsourced data and the WiFi-Direct discovery scans: (a) $\langle 1\,hop, 5\,min\rangle$, (b) $\langle 4\,hops, 5\,min\rangle$ (default configuration), and (c) $\langle 4\,hops, 10\,min\rangle$;

3- FOUGERE *with LPPMs* refers to the FOUGERE library with the default configuration 2-b selected with 2 privacy distortions—*location noise* and *time noise*—and 1 privacy aggregation—*k-anonymity*, which are representative LPPMs used by the state-of-the-art. To configure these LPPMs, we consider 2 worker profiles, which are mapped to the following values:
   (a) *weak privacy profile* where location noise is set to $\langle 1, 0.1, 0.05\rangle$, thus reducing the location precision by 1 digit with a probability of 0.1 and possibly removing the location with a probability of 0.05. Time noise is set to $\langle 30, 0.1, 0.05\rangle$, thus reducing the time precision to half an hour with a probability of 0.1 and possibly removing the timestamp with a probability of 0.5, and finally k-anonymity is set to $\langle 2\rangle$, meaning that at least 2 samples should be produced in the same area to be forwarded;
   (b) *strong privacy profile* configured with location noise = $\langle 2, 0.2, 0.1\rangle$, time noise = $\langle 60, 0.2, 0.1\rangle$ and k-anonymity = $\langle 4\rangle$ as privacy settings.

None of these configurations includes a privacy filter as these LPPMs are expected to be used to hide the living and working places of workers and the input dataset does not include this information. Furthermore, this paper does not

aim at evaluating the efficiency of individual LPPMs, but rather demonstrating the benefit of combining them in an open framework like FOUGERE.

Performance analysis. FOUGERE implements an opportunistic dissemination scheme to improve the privacy of workers. By doing so, FOUGERE exploits the physical proximity of workers to exchange crowdsourced data and to guarantee that the uploaded data has been forwarded along a number of hops requested by the worker. Figure 3 depicts the *time to converge* as a metrics to evaluate *i)* the impact of integrating FOUGERE on a legacy mobile crowdsourcing app like MOBIPERF, and *ii)* the effect of the number of hops and the WiFi-Direct discovery duration parameters. One can observe that, by using FOUGERE, not all the crowdsourced data is reported back to the remote storage server. This can be explained by the fact that some workers are contributing in sparsely populated areas, which prevents FOUGERE from disseminating the collected measurements. This result is actually a strength of FOUGERE as it automatically protects the workers from adversaries who would apply some location distribution attacks to identify them.

Regarding the parameters of FOUGERE, one can note that the delay to upload data and the volume of reported data is more affected by the discovery duration than the number of hops required to upload the crowdsourced data. By increasing the delay of peer discovery, mobile devices miss some other workers in their vicinity in order to improve the time to converge. Therefore, we privilege the configuration 2-b (4 hops and 5 minutes) as the default configuration for FOUGERE. However, the worker remains free to adjust each of these parameters.

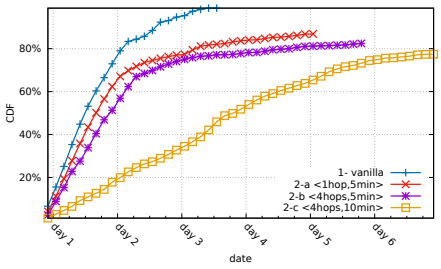

Fig. 3. Measurements' time to converge

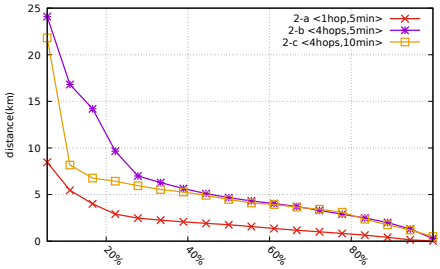

Fig. 4. Distance traveled by measurements

The *traveling distance* is another interesting metrics to evaluate the efficiency of the dissemination process and the relevance of peer-to-peer communications. Increasing this data traveling distance with FOUGERE contributes to better shuffle crowdsourced data produced by a crowd of workers. Figure 4 reports on this distance traveled by the crowdsourced data before being uploaded back to the remote storage server. In particular, the default configuration of FOUGERE maximizes the traveled distance with 20 % of data that traveled at least 10 km

(6.2 *miles*), thus ensuring that the data was conveyed by FOUGERE as far as possible from the location where it has been produced.

Utility analysis. While FOUGERE aims at improving the location privacy of workers, the utility of the resulting dataset should not be neglected. Figure 5 reports on the tradeoff between utility and anonymity of the configurations we considered. While the vanilla configuration (1) offers the highest utility with no anonymity, one can observe that the integration of FOUGERE seriously improves the anonymity of workers without seriously impacting the utility of the resulting dataset. As mentioned in Figure 3, the loss of 20 % utility is mainly due to crowdsourced data in sparsely populated areas that were retained by FOUGERE. Furthermore, adding some LPPMs (configurations 3-a and 3-b) strongly increase the anonymity of workers.

Interestingly, one can observe that the *weak privacy profile* offers a good privacy/anonymity tradeoff compared to the *strong privacy profile*, which seriously harms the dataset utility without bringing any further improvement over anonymity.

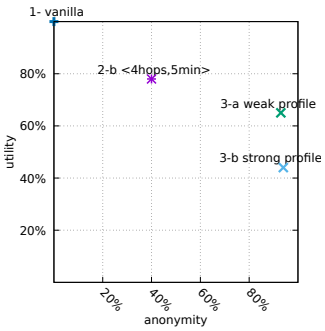

Fig. 5. Dataset utility

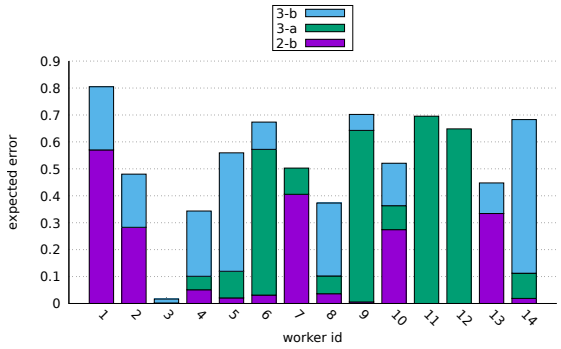

Fig. 6. Robustness against location privacy attacks

Robustness analysis. Regarding the effective privacy support offered by FOUGERE, we used the LPM^2 toolkit to evaluate the robustness of crowdsourced datasets that are uploaded through FOUGERE. We randomly select the crowdsourced data reported by one of the workers as the adversary knowledge required by LPM^2 to apply location privacy attacks and we depict in Figure 6 the reported robustness for 14 workers. While LPM^2 successfully defeats worker 3 (used as the adversary knowledge), the other 14 workers clearly benefit from the integration of FOUGERE. In particular, we can observe that the integration of LPPMs complements efficiently our opportunistic dissemination scheme by supporting workers who are not located in a dense area and by offering similar privacy guarantees. Successfully location privacy attacks requires to combine different strategies to cope with the profile of workers.

While FOUGERE offers the worker the possibility to manually adjust her privacy settings, one of the perspectives of this work consists in leveraging this configuration process by delivering privacy risk feedback that would guide her settings accordingly. By recommending the privacy settings of FOUGERE, we aim at maximizing the individual privacy of workers, while preserving the overall utility of the crowdsourced dataset (cf. Figure 5).

Uncertainty analysis. Finally, we consider the view of an adversary to study the level of uncertainty that introduces FOUGERE into the crowdsourced dataset. Figure 7 reports on the uncertainty metrics computed by LPM^2. One can observe that FOUGERE succeeds to increase the uncertainty of adversaries when it combines the opportunistic dissemination scheme with LPPMs, which confirms our previous observation. Furthermore, it also assesses that adopting a *weak privacy profile* already brings a reasonable level of privacy that puts adversaries in difficulties.

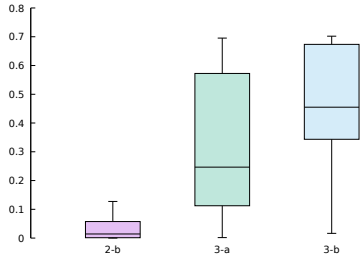

Fig. 7. Adversary uncertainty

Indicator	**Value**
Crowdsourced dataset size	29,712
Exchanged messages	113,785
Contributions per user	59
Messages forwarded per user	227
Detected neighbors	1,730,827
Established connections	127,545
Isolated users	8

Fig. 8. Overhead analysis for 500 workers

Overhead analysis. To analyze the overhead induced by our data dissemination process, we report in Table 8 the statistics related to an experiment involving 500 emulated workers for 24 hours. Along the experiment, the workers adopt the default configuration of FOUGERE $\langle 4\,hops, 5\,min\rangle$ (2-b). The overhead per user and at the scale of the crowd does not exceed 4 times the initial volume of contributions. FOUGERE also discards 8 users considered as isolated and thus identifiable by tools like LPM^2.

7 Threats to Validity

This section analyzes the factors that may threaten the validity of our results.

Internal validity concerns the relation between theory and observations. In this paper, they could be due to measurement errors reported during the experimentations. That is the reason why we did several experiments and we tried to reduce as much as possible external factors as explained in our experimental

protocol in Section 6.2. We also performed our experiments on a crowd of emulated devices equipped with real mobile apps, instead of a simulation, to reduce the threats that could be due to an integration of the proposed approach in a real mobile crowdsourcing app or platform.

External validity relates to the possibility to generalize our findings. We believe that further validations should be done on different mobile crowdsourcing apps and with different configurations to broaden our understanding of the impact of LPPMs on the privacy of workers. Thus, we are not assuming that our results can be used to generalize the impact of a specific LPPM on privacy. However, we believe that this paper contributes to prove that there is a clear positive impact for the privacy threats we considered.

Reliability validity focuses on the possibility of replicating our experiments and results. We attempt to provide all the necessary details to replicate our study and our analysis. Furthermore, the reference implementation of FOUGERE, the input datasets, case studies and testing environment are made available online to leverage its reproduction by the research community.

Construct validity has been covered by considering the convergent validity of privacy and utility properties. We observed that these two properties are related in practice, as the application of LPPMs tends to decrease the utility of the crowdsourced dataset. This observation calls for the identification of a privacy and utility trade-off in the context of mobile crowdsourcing systems, as acknowledged by [5].

Conclusion validity refers to the correctness of the conclusions reached in this paper. The empirical evaluation we reported confirms our initial assumption that *a priori* anonymization techniques can be used to leverage the privacy of workers. We were also careful with our conclusion with regards to the impact on the utility of crowdsourced dataset.

8 Conclusion

Mobile crowdsourcing apps and platforms are more and more challenged to protect their workers' privacy. To address this challenge, we introduce FOUGERE to increase worker's privacy in mobile crowdsourcing systems. FOUGERE operates a system-level service that collaborates with a mobile crowdsourcing app to declare SPI and delegate the dissemination of crowdsourced data by leveraging the physical proximity of workers. This opportunistic dissemination scheme is complemented by the integration of LPPMs that can be configured by the workers, independently of the installed mobile crowdsourcing apps.

Finally, we consider the deployment of FOUGERE in a realistic Android environment by emulating a crowd of 15 mobile devices hosting different versions of MOBIPERF and FOUGERE to assess our contribution. We show that FOUGERE succeeds to improve the workers' privacy by defeating location privacy attacks implemented by the LPM^2 toolkit.

References

1. Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C.: Geo-indistinguishability: differential privacy for location-based systems. In: Proc. of CCS'13. pp. 901–914 (2013)
2. Balan, R.K., Misra, A., Lee, Y.: LiveLabs: Building An In-Situ Real-Time Mobile Experimentation Testbed. In: ACM HotMobile (2014)
3. Boutsis, I., Kalogeraki, V.: Location privacy for crowdsourcing applications. In: Proc of. UbiComp'16 (2016)
4. Brouwers, N., Langendoen, K.: Pogo, a Middleware for Mobile Phone Sensing. Proc. of Middleware'12 (2012)
5. Cerf, S., Primault, V., Boutet, A., Mokhtar, S.B., Birke, R., Bouchenak, S., Chen, L.Y., Marchand, N., Robu, B.: PULP: Achieving Privacy and Utility Trade-off in User Mobility Data. In: Proc. of SRDS'17 (Sep 2017)
6. Chatzimilioudis, G., Konstantinidis, A., Laoudias, C., Zeinalipour-Yazti, D.: Crowdsourcing with smartphones. IEEE Internet Computing **16**(5) (sep 2012)
7. Chen, R., Fung, B.C.M., Mohammed, N., Desai, B.C., Wang, K.: Privacy-preserving trajectory data publishing by local suppression. Information Sciences **231** (2013)
8. Choi, H., Chakraborty, S., Charbiwala, Z.M., Srivastava, M.B.: SensorSafe: A framework for privacy-preserving management of personal sensory information. In: Proc. of SDM'11. vol. 6933 LNCS (2011)
9. Chow, C.Y., Mokbel, M.F., Liu, X.: A peer-to-peer spatial cloaking algorithm for anonymous location-based service. Proc. of ACM SIGSPATIAL (2006)
10. Chow, C.Y., Mokbel, M.F., Liu, X.: Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. GeoInformatica **15**(2) (apr 2011)
11. Christin, D., Bub, D.M., Moerov, A., Kasem-Madani, S.: A distributed privacy-preserving mechanism for mobile urban sensing applications. In: Proc of. ISSNIP'15 (apr 2015)
12. Christin, D., Reinhardt, A., Kanhere, S.S., Hollick, M.: A survey on privacy in mobile participatory sensing applications. Journal of Systems and Software **84**(11) (2011)
13. Cornelius, C., Kapadia, A., Kotz, D., Peebles, D., Shin, M., Triandopoulos, N.: Anonysense: privacy-aware people-centric sensing. Proc. of Mobisys'08 (2008)
14. Das, T., Mohan, P., Padmanabhan, V.N., Ramjee, R., Sharma, A.: PRISM: Platform for Remote Sensing using Smartphones. Proc. of MobiSys'10 (2010)
15. Fawaz, K., Shin, K.G.: Location privacy protection for smartphone users. In: Proc. of CCS'14. ACM (2014)
16. Gambs, S., Killijian, M.O., Cortez, M.N.d.P.: GEPETO: A GEoPrivacy-Enhancing TOolkit. In: Proc. of AINA Workshops'10 (2010)
17. Gambs, S., Killijian, M.O., Del Prado Cortez, M.N.: Next place prediction using mobility Markov chains. In: Proc. of MPM'12 (2012)
18. Gambs, S., Killijian, M.O., Núñez del Prado Cortez, M.: De-anonymization attack on geolocated data. Journal of Computer and System Sciences **80**(8) (2014)
19. Gao, S., Ma, J., Shi, W., Zhan, G., Sun, C.: TrPF: A trajectory privacy-preserving framework for participatory sensing. IEEE Transactions on Information Forensics and Security **8**(6) (jun 2013)
20. Haderer, N., Rouvoy, R., Seinturier, L.: A preliminary investigation of user incentives to leverage crowdsensing activities. Proc. of PerCom'13 (2013)

21. Hu, L., Shahabi, C.: Privacy assurance in mobile sensing networks: Go beyond trusted servers. In: 2010 8th IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops). pp. 613–619. IEEE (2010)
22. Huang, J., Chen, C., Pei, Y., Wang, Z., Qian, Z., Qian, F., Tiwana, B., Xu, Q., Mao, Z., Zhang, M., Others: Mobiperf: Mobile network measurement system. Tech. Report. University of Michigan and Microsoft Research (2011)
23. Kifer, D.: l-Diversity : Privacy Beyond k -Anonymity. Proc. of ICDE'06 **1**(1) (mar 2006)
24. Krumm, J.: Inference Attacks on Location Tracks. Pervasive Computing **10**(Pervasive) (2007)
25. Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J.I., Zhang, J.: Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In: Proc. of UbiComp'12 (2012)
26. Ma, C.Y., Yau, D.K., Yip, N.K., Rao, N.S.: Privacy vulnerability of published anonymous mobility traces. In: Proc. of MobiCom'10 (2010)
27. Meftah, L., Gomez, M., Rouvoy, R., Chrisment, I.: ANDROFLEET: Testing WiFi Peer-to-Peer Mobile Apps in the Large. Proc. of ASE'17 (2017)
28. Ninghui, L., Tiancheng, L., Venkatasubramanian, S.: t-Closeness: Privacy beyond k-anonymity and l-diversity. In: Proc. of ICDE'07 (2007)
29. Peng, T., Liu, Q., Meng, D., Wang, G.: Collaborative trajectory privacy preserving scheme in location-based services. Information Sciences **387** (2017)
30. Piorkowski, M., Sarafijanovic-Djukic, N., Grossglauser, M.: CRAWDAD dataset epfl/mobility (v. 2009-02-24). Downloaded from `https://crawdad.org/epfl/mobility/20090224` (Feb 2009). https://doi.org/10.15783/C7J010
31. Prandi, C., Salomoni, P., Mirri, S.: mPASS: Integrating People Sensing and Crowdsourcing to Map Urban Accessibility. Proc. of CCNC'14 (2014)
32. Shokri, R., Theodorakopoulos, G., Le Boudec, J.Y., Hubaux, J.P.: Quantifying location privacy. In: Proc. of S&P'11 (may 2011)
33. Shokri, R., Theodorakopoulos, G., Papadimitratos, P., Kazemi, E., Hubaux, J.P.: Hiding in the mobile crowd: Location privacy through collaboration. IEEE Transactions on Dependable and Secure Computing **11**(3) (may 2014)
34. Sweeney, L.: k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. International Journal on Uncertainty **10**(5) (2002)
35. Terrovitis, M., Mamoulis, N.: Privacy preservation in the publication of trajectories. Proc. of MDM'08 (2008)
36. Verykios, V.S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., Theodoridis, Y.: State-of-the-art in privacy preserving data mining. ACM SIGMOD Record **33**(1) (2004)
37. Wernke, M., Skvortsov, P., Dürr, F., Rothermel, K.: A classification of location privacy attacks and approaches. Personal Ubiquitous Comput. **18**(1) (Jan 2014). https://doi.org/10.1007/s00779-012-0633-z