Hiding in the Mobile Crowd: Location Privacy through Collaboration

Reza Shokri, George Theodorakopoulos, Panos Papadimitratos, Ehsan Kazemi,
Jean-Pierre Hubaux, *Fellow, IEEE*

Abstract—Location-aware smartphones support various location-based services (LBSs): users query the LBS server and learn on the fly about their surroundings. However, such queries give away private information, enabling the LBS to track users. We address this problem by proposing the first user-collaborative privacy-preserving approach for LBSs. Our solution does not require changing the LBS server architecture, and it does not assume third party servers; still, it significantly improves users' location privacy. The gain stems from the collaboration of mobile devices: they keep their context information in a buffer and pass it to others seeking such information. Thus, a user remains hidden from the server, unless all the collaborative peers in the vicinity lack the sought information. We evaluate our scheme against Bayesian localization attacks, which allow for strong adversaries who can incorporate prior knowledge in their attacks. We develop a novel epidemic model to capture the, possibly time-dependent, dynamics of information propagation among users. Used in the Bayesian inference framework, this model helps analyze the effects of various parameters, such as the users' querying rate and the lifetime of context information, on users' location privacy. The results show that our scheme hides a high fraction of location-based queries, thus significantly enhancing users' location privacy. Our simulations with real mobility traces corroborate our model-based findings. Finally, our implementation on mobile platforms indicates that it is lightweight and the collaboration cost is negligible.

Index Terms—Mobile Networks, Location-based Services, Location Privacy, Bayesian Inference Attacks, Epidemic Models

1 Introduction

Smartphones, among other increasingly powerful mobile computing devices, offer various methods of localization. Integrated GPS receivers or positioning services based on nearby communication infrastructure (WiFi access points or base stations of cellular networks) enable users to position themselves fairly accurately, which has led to a wide offering of *Location-Based Services* (LBSs). Such services can be queried by users to provide real-time information related to the current position and surroundings of the device, e.g. contextual data about points of interest such as petrol stations, or more dynamic information such as traffic conditions. The value of LBSs is exactly in obtaining accurate and up-to-date information on the fly.

Convenient though LBSs may be, disclosing location information can be dangerous. Each time an LBS query is submitted, private information is revealed. The user can be linked to her location, and multiple pieces of such information can be linked together. Users can then be profiled, leading to unsolicited targeted advertisements or price discrimination.

Even worse, from a person's whereabouts one can infer her habits, personal and private preferences, religious beliefs, political affiliations, etc. That could make her the target of blackmail or harassment. Finally, real-time location disclosure leaves a person vulnerable to absence disclosure attacks: learning that someone is away from her home could allow a house break-in or blackmail [1]. Knowing the real-time location of a person could lead to stalking.

All this information is collected by the LBS operators. So, they may be tempted to misuse the rich data they collect by e.g. selling it to advertisers or to private investigators. The mere existence of such valuable data may invite attackers, who could break into the LBS servers and obtain logs of user queries, or governments that want to detect and suppress dissident behavior. The result in all cases is the same: user-sensitive data fall in the hands of untrusted parties.

The difficulty of the problem lies in protecting user privacy while continuing to reap the benefits of LBSs. Therefore, solutions such as not using LBSs are not acceptable. For instance, a user could download a large volume of data and then search through it for specific context information as the need arises. But this would be cumbersome, if not impractical, and it would be inefficient for obtaining information that changes dynamically over time.

The need to enhance privacy for LBS users has been understood and several solutions have been proposed, falling roughly into two main categories: centralized and user-centric.

- *R. Shokri is with ETH Zurich, Switzerland.*
  E-mail: reza.shokri@inf.ethz.ch
- *G. Theodorakopoulos is with Cardiff University, UK.*
  Email: g.theodorakopoulos@cs.cardiff.ac.uk
- *P. Papadimitratos is with KTH, Stockholm, Sweden.*
  Email: papadim@kth.se
- *E. Kazemi and J.-P. Hubaux are with EPFL, Switzerland.*
  E-mail: firstname.lastname@epfl.ch

Centralized approaches introduce a third party in the system that protects users' privacy, operating between the user and the LBS. Such an intermediary proxy server could anonymize (and obfuscate) queries by removing any information that identifies the user or her device. Alternatively, it could blend a user's query with those of other users, so that the LBS server always sees a group of queries [2]. However, such approaches only shift the problem: the threat of an untrustworthy LBS server is addressed by the introduction of a new third-party server. Why would the new server be any more trustworthy? Additionally, new proxy servers become as attractive for attackers as centralized LBSs.

Other centralized approaches require the LBS to change its operation by, for example, mandating that it process modified queries (submitted in forms that are different from actual user queries, possibly encrypted using PIR techniques [3]), or that it store data differently (e.g., encrypted or encoded, to allow private access [4]). Centralized interventions or substantial changes to the LBS operation would be hard to adopt, simply because the LBS providers would have little incentive to fundamentally change their operation. Indeed, if a revenue stream is to be lost by not collecting user data, not many LBS providers can be expected to comply. Misaligned incentives have been identified as the root of many security problems [5].

User-centric approaches operate on the device. Typically they aim to blur the location information by, for example, having the user's smartphone submit inaccurate, noisy GPS coordinates to the LBS server. However, obfuscation approaches (e.g., spatial/temporal cloaking) that protect user location-privacy can degrade the user experience if users need high privacy: e.g., LBS responses would be inaccurate or untimely. Obfuscation also is not effective against absence disclosure [6].

Our approach avoids the problems of these two extremes by having users collaborate with each other to jointly improve their privacy, without the need for a trusted third-party (TTP). In effect, the mobile crowd acts as a TTP, and the protection mechanism becomes a distributed protocol among users. Mobile users concerned about their location privacy are indeed the most motivated entities to engage in protecting themselves. We require no change of the LBS server architecture and its normal operation, and we make no assumption on the trustworthiness of the LBS or any third-party server.

The key idea of our scheme, MobiCrowd, is that users only contact the LBS server if they cannot find the sought information among their peers, i.e., other nearby reachable user devices. Thus, users can minimize their location information leakage by *hiding* in the crowd. Clearly, MobiCrowd would be most effective when there are many peers gathered at the same location. Indeed, this clustering phenomenon has been observed in human mobility studies [7]. Moreover, the places where people gather are points of interest, where users are most likely to query an LBS. Thus, MobiCrowd would be used exactly where it is most effective.

We evaluate MobiCrowd through both an *epidemic-based differential equation model* and a *Bayesian framework for location inference attacks*. The epidemic model is a novel approach to evaluating a distributed location-privacy protocol. It helps us analyze how the parameters of our scheme, combined with a time-dependent model of the users' mobility, could cause a high or low degree privacy. We validate the model-based results (on the probability of hiding a user from the server) with simulations on real mobility traces. We find that our epidemic model is a very good approximation of the real protocol, reflecting the precise hiding probability of a user, in various settings.

Relying on hidden Markov models, the Bayesian inference framework quantifies the correctness with which an adversary can estimate the location of users over time. The error of the adversary in this estimation is exactly our privacy metric [8]. We evaluate MobiCrowd on a real location dataset and we show that it provides a high level of privacy for users with different mobility patterns, against an adversary with varying background knowledge.

Note that this joint epidemic/Bayesian evaluation is necessary and, in fact, a significant component of our approach, as MobiCrowd is a *distributed protocol* running on multiple collaborating devices, so its performance depends on network characteristics (e.g. time-dependent mobility), not just on what an individual device does. The focus of the existing work in the literature is more on privacy-preserving *functions* (e.g., obfuscation functions run independently by each user [9], [10]). To the best of our knowledge, this is the first such evaluation, and it is significantly more realistic than our own previous work [11] that quantified privacy with just the fraction of queries hidden from the server.

We have implemented our scheme on the Nokia N800, N810 and N900 mobile devices, and demonstrated it with the Maemo Mapper (a geographical mapping software for points of interest) [12]. Our approach can be ported to the upcoming technologies that enable mobile devices to directly communicate to each other via (more energy-efficient) Wi-Fi-based technologies [13], [14] that aim at constructing a mobile social network between mobile users.

The rest of the paper is organized as follows. We survey the related work in Section 2. In Section 3, we describe our assumptions for the Location-Based Service, for the mobile users and the adversary, and we state our design objectives. We present MobiCrowd in Section 4, and then we develop an epidemic model of its operation in Section 5. We present our Bayesian localization attacks in Section 6. We evaluate the effectiveness of MobiCrowd in Section 7. We conclude the paper in Section 8.

2 Related Work

There are many collaborative schemes for mobile networks. Mobile users, for example, can collectively build a map of an area [15]. Collaboration is also needed when sharing content or resources (e.g. Internet access) with other mobile nodes [16].

Various threats associated with sharing location information have been identified in the literature. For example, users can be identified even if they share their location sporadically [17]. Social relations between users can help an adversary to better de-anonymize their location traces [18]. Finally, location sharing of a user does not only diminish her own privacy, but also the privacy of others [19].

Techniques proposed to protect location privacy in LBSs can be classified based on how they distort the users' queries before the queries reach the LBS server. The queries can be *anonymized* (by removing users' identities), *pseudonymized* (by replacing users' real names with temporal identifiers called pseudonyms), or *obfuscated* (by generalizing or perturbing the spatiotemporal information associated to the queries). Queries can also be camouflaged by adding some *dummy queries*, or be completely eliminated and *hidden* from the LBS [10]. Combinations of these methods have been employed in the existing (centralized or distributed) mechanisms. We now discuss these approaches in more detail.

The mere anonymization of (especially the continuous) queries does not protect users' location privacy: the queries of a user are correlated in space and time; hence, the adversary can successfully link them to each other by using target tracking algorithms [20] or can successfully identify the real names of the users [21]. Changing user pseudonyms while the users are passing through pre-defined spots, called mix zones [22], makes it difficult to track the users along their trajectories. However, users must remain silent inside the mix zones, which means that they cannot use the LBS. To mitigate this problem, the size of the mix zones is kept small, which in turn limits the unlinkability of users' queries. Even if the mix zones are optimally placed, the adversary's success is relatively high [23].

Perturbing the query's spatiotemporal content, in addition to anonymization by a third party (central anonymity server), has been proposed for obtaining a higher level of privacy [2]. The main drawback is the reliance on a centralized third party, limiting the practicality of this proposal. The considerable degradation of the quality of service imposed by obfuscation methods is another deterrent for such solutions. In [24], for example, the need to construct the cloaking regions and also to receive the responses from the server through other users can considerably degrade the service. Many obfuscation-based techniques are based on k-anonymity, which has been shown inadequate to protect privacy [8], [25]. Perturbation techniques with differential privacy guarantee, however, have been shown effective against an adversary with arbitrary knowledge [26].

Adding dummy queries to the user actual queries might help to confuse the adversary about the actual user location. But generating effective dummy queries that divert the adversary is a difficult task [27], as they need to look like actual queries over space and time. An optimum algorithm for generating dummy queries is an open problem.

In all the above-mentioned mechanisms, there is always a trade-off between users' privacy and the quality of service they experience [28]. The tension is maximized when it comes to *hiding* queries from the LBS server. Hiding a query from the server minimizes the revealed user information and thus maximizes her privacy with respect to that query. Simply put, it is more effective than the other three privacy protection methods, and it protects users against both presence and absence disclosure. This is what MobiCrowd provides: Hiding from the server while receiving the query responses from other peers.

There exist cryptographic approaches that redesign the LBS: the service operator does not learn much about the users' queries, while it can still reply to their queries [4], or it can obtain imprecise information about user location [3]. The lack of incentives for LBS operators to change their business model and implement these solutions, and their high computational overhead have made them impractical so far.

A game-theoretic evaluation of our protocol run by rational users is presented in [29].

3 Problem Statement

3.1 Mobile Users and LBS

We consider N users who move in an area split into M discrete regions/locations. The mobility of each user u is a discrete-time Markov chain on the set of regions: The probability that user u, currently in region r_i, will next visit region r_j is denoted by $p_u(r_j|r_i)$. Let $\pi_u(r_i)$ be the probability that user u is in region r_i.

Each user possesses a location-aware wireless device, capable of ad hoc device-to-device communication and of connecting to the wireless infrastructure (e.g., cellular and Wi-Fi networks). As users move between regions, they leverage the infrastructure to submit local-search queries to an LBS, at some frequency that we term LBS access frequency. The frequency at which users query the LBS varies depending on the type of requested information, the dynamics of information update in the LBS database, or the geographical region.

The information that the LBS provides expires periodically, in the sense that it is no longer valid. Note that information expiration is not equivalent to the user accessing the LBS: A user accesses the LBS when her information has expired *and* she wishes to receive the most up-to-date version of it.

In addition, the information the LBS provides is *self-verifiable*, i.e., users can verify the integrity and authenticity of the server responses. This can be done in different ways; in our system, the user device verifies a digital signature of the LBS on each reply using the LBS provider's public key. As a result, a compromised access point or mobile device cannot degrade the experience of users by altering replies or disseminating expired information.

3.2 Adversary Model and Privacy Metric

LBS servers concentrate location information from all user queries. Thus, an untrusted service provider could act as a "big brother," that is, it could monitor user whereabouts and activities over time. In such a setting, the adversary can be categorized as a *passive global long-term* observer [10]. We assume the adversary has some *background knowledge* about the users' mobility patterns. This background knowledge consists of each user's mobility model, expressed as a Markov chain, the users' LBS access frequency, and the information lifetime.

The adversary aims to perform inference attacks against the locations of users. In other words, he uses his background knowledge to estimate the locations from which the users issue queries, but also the locations they visit between successive queries that are not directly disclosed to the LBS.

We *quantify* the location privacy of users as the expected error of the adversary in estimating the actual location of each user at each time instant [30]. The more queries the adversary observes, the more successful he will be in reconstructing their actual trajectories; so privacy is proportional to the distortion of the reconstructed trajectories.

We do not address the threat of local observers sniffing the wireless channel trying to infer users' private information, as such a threat could exist with or without MobiCrowd, and it can be alleviated by frequently changing device identifiers (e.g., changing MAC addresses for WiFi networks [31] similar to changing TMSI for GSM networks [32]). More importantly, local observers, to be effective, would need to be *physically* present next to any given victim user, over long periods and across different locations. In contrast, a centralized LBS can by default observe *all* the queries of a user, which is why we focus on this much greater threat in this paper.

Malicious users cannot mislead others into receiving fake information, because messages are digitally signed by the LBS (as assumed in the previous subsection).

3.3 Design Objectives

Overall, we seek to design a practical and highly effective location-privacy preserving mechanism for LBSs: We should protect privacy with a minimal compromise on LBS quality of service. The nature of existing threats and the structure of stakeholder incentives, outlined earlier, is the determining factor of our design objectives.

Our first design objective is to ***not*** *rely on architectural changes of the LBS*; any such changes would be impractical and highly unlikely to be adopted. Relying on centralized trusted third parties (e.g., central anonymity servers) to provide privacy enhancing mechanisms can be as hard as having trusted LBS operators. This leads to our second design objective: *no reliance on any third party server to provide privacy protection*. In fact, we would like to *place the privacy protection* exactly where there is incentive and motivation, that is, *with the users themselves*.

4 Our Scheme

Based on the stated design objectives, we propose a novel location-privacy preserving mechanism for LBSs. To take advantage of the high effectiveness of hiding user queries from the server, which minimizes the exposed information about the users' location to the server, we propose a mechanism in which a user can ***hide*** *in the mobile crowd* while using the service.

The rationale behind our scheme is that users who already have some location-specific information (originally given by the service provider) can pass it to other users who are seeking such information. They can do so in a wireless peer-to-peer manner. Simply put, information about a location can "remain" around the location it relates to and change hands several times before it expires. Our proposed collaborative scheme enables many users to get such location-specific information from each other *without contacting the server*, hence minimizing the disclosure of their location information to the adversary.

4.1 Scheme Details

We build a *mobile transparent proxy* in each device that maintains a buffer with location-specific information. This buffer keeps the replies the user obtains from the server or other peers. Each piece of information associated with a given region has an expiration time (which is attached to the information and protected with the digital signature), after which the information is no longer valid. Invalid information is removed from the buffer.

Each user with valid information about a region is termed *informed user* for that region. Users interested in getting location-specific information about a region are called information *seekers* of that region. A seeker, essentially a user who does not have the sought information in her buffer, first broadcasts her query to her neighbors through the wireless ad hoc interface of the device. We term this a *local query*.

Any of the receivers of such a local query may respond to it, by what we term a *local reply*, as long as it has the information its peer seeks. However, an informed device will not necessarily respond to any received query: this will happen if the device is not only informed, but also *willing to collaborate*. We design our system with this option for its users; the collaborative status may be set explicitly by the user or automatically recommended or set by the device. Simply put, having each user collaborate a limited number of times (a fraction of the times she receives a local query from her neighbors), or during a randomly selected fraction of time, balances the cost of collaboration with the benefit of helping other peers. In practice, this is equivalent to the case where only a fraction of users collaborate.

By obtaining a local reply, the seeker is now informed while, more importantly, her query has remained hidden from the service provider. No privacy-sensitive information has been exposed to the server and the user has obtained the sought service. Of course, in case there is no informed user around the seeker willing to assist her, she has no choice but to contact the server directly. In essence, a subset of users in every region have to contact the LBS to get the updated information, and the rest of the users benefit from the peer-to-peer collaboration. Intuitively, the higher the proportion of hidden user queries, the higher her location privacy will be.

5 Epidemic Model for the Dynamics of MobiCrowd

The performance of our system depends on various parameters, such as the rate of contacts and the level of collaboration between users, the rate of LBS query generation, etc. We now describe a model for MobiCrowd, with the help of which we can directly evaluate the effect of various parameters on users' location privacy. Observing the effect of the parameters also helps when designing a system and testing "what-if" scenarios. For example, we can immediately see the level of collaboration required to achieve a desired privacy level or how the privacy level will change if the users make queries more frequently or less frequently.

We draw an analogy between our system and *epidemic phenomena*: location-context information spreads as an infection from one user to another, depending on the user state (seeking information, having valid information, etc.). For example, a seeker becomes "infected" when meeting an "infected" user, that is, a user with valid information.

We want a model that describes transitions between, and keeps track of, the various states a user is in as time progresses. However, it is prohibitively complex to keep track of the state of each individual user. Therefore, we make use of the *mean field approximation* [33], which focuses on the fraction of users in each state; these fractions are collectively called the *network state*. The approximation applies when the number of users is large and each individual interaction contributes a vanishingly small change to the network state. The approximation requires a random contact pattern among users, rather than a spatially correlated pattern, and random contacts are not far from reality when users are clustered in the same region (recall that we partition the whole area into regions).

The mean field approximation tells us that the time evolution of the fraction of users in each state can be described with increasing accuracy, as the number of users grows, by a system of Ordinary Differential Equations (ODEs). By studying the system of ODEs, we find the steady state(s) to which the network converges. Similar models have been used in human virus epidemics [34], in worm propagation in wireless networks [35], and in research on forwarding/gossiping protocols [36].

To keep the presentation simple we focus on one type of context information, hence we consider a single average information lifetime. No loss of generality results from this, because, to model a complete system with multiple types of information, we can merge multiple versions of this model, one for each type.

5.1 Model States and System of ODEs

As mentioned earlier, users move in an area partitioned into multiple regions. The state of context knowledge within a region intuitively corresponds to the disease status in an epidemic. In general, a user's knowledge state would be multi-dimensional, because a different piece of information is relevant for each region. Thus, for each region we would have an associated epidemic model, with the same structure but different parameters. However, the state of knowledge about a region is unrelated to the knowledge about other regions, so different regions can be analyzed separately. We present our model for a single region, with users entering and exiting it, and we describe the states and the dynamics of our epidemic model for that single region.

The collective mobility of users with respect to a region is modeled using three parameters: β, the average number of times a user makes a proximity contact with other users per time unit within a region; μ, the probability of an outsider user enters a region within a time unit; and λ, the probability of an insider user leaves a region within a time unit. We derive these parameters from the Markov mobility models of users, as follows. Let parameters λ_i and μ_i be the probabilities of exiting and entering region r_i, respectively. They correspond to the expected number of users who exit/enter r_i normalized by the expected number of users who are inside/outside of r_i.

$$\lambda_i = \frac{\sum_{u,j\neq i} \pi_u(r_i) p_u(r_j|r_i)}{\sum_u \pi_u(r_i)} \tag{1}$$

$$\mu_i = \frac{\sum_{u,j\neq i} \pi_u(r_j) p_u(r_i|r_j)}{\sum_u (1-\pi_u(r_i))} \tag{2}$$

The contact rate β_i between users in region r_i corresponds to the expected number of contacts of a device within its communication range.

$$\beta_i = \sum_{k=0}^{n_i-1} k \binom{n_i-1}{k} q^k (1-q)^{n_i-1-k} \tag{3}$$

where q is the fraction of region's area that is within the user's communication range, and $n_i = \sum_u \pi_u(r_i)$ is the expected number of users in region r_i. Note that the mobility parameters (λ, μ, and β) can also be computed directly from sample location traces. The list of all parameters of the epidemic model are listed in Table 1.

Seeker: Users who are *interested in obtaining information* (i.e., have requested the information but not yet received it) are in the Seeker state. Once they have it, they move into the Informed state. As long as a Seeker user stays in the region that she seeks information about, she is called an *Insider Seeker*. These users can receive information from other Informed users in the region, or from the server, the ultimate source of information. A Seeker who leaves the region after requesting information about that region is called an *Outsider Seeker*. An Outsider Seeker can only receive information from the server, as users need to be in the same region in order to be able to propagate information to each other.

Informed: Users who *have information* about the region are in the Informed state. If they are *inside the region* (called *Insider Informed*), they accept to spread the information at each contact with a Seeker user with probability ϕ. This is because the information spreading process imposes some communication cost on Informed users and, hence, they might not always collaborate. If they are *outside the region* (called *Outsider Informed*), we assume they do not spread the information. The information that the Informed users have, whether they are inside or outside the region, expires with rate δ and the users become Removed.

Removed: Users who *do not have information* and are *not currently interested in obtaining information* are in the Removed state. We distinguish between Insider Removed and Outsider Removed users. An Insider Removed user becomes a Seeker if the user becomes interested in obtaining information about the region. As LBS users usually query information about the region they are in, we assume that outsiders have to enter the region to become interested.

$S(t)$	insider Seeker users at time t
$S^*(t)$	outsider Seeker users at time t
$I(t)$	insider Informed users at time t
$I^*(t)$	outsider Informed users at time t
$R(t)$	insider Removed users at time t
$R^*(t)$	outsider Removed users at time t
λ	probability of exiting the region within a time unit
μ	probability of entering the region within a time unit
β	contact rate per user per time unit
γ	avg request rate per user per time unit
$1/\omega$	avg waiting time before contacting the server
$1/\delta$	information avg lifetime
ϕ	avg collaboration probability

TABLE 1
List of the symbols used in the epidemic model

We denote by $S(t)$, $S^*(t)$, $I(t)$, $I^*(t)$, $R(t)$, and $R^*(t)$, respectively, the fraction of Seeker Insider, Seeker Outsider, Informed Insider, Informed Outsider, Removed Insider, and Removed Outsider users of a given region at time t. The *network state* $y(t)$ is the vector of these values. The system of equations that models the evolution of the network state is

$$S(t) + S^*(t) + I(t) + I^*(t) + R(t) + R^*(t) = 1 \tag{4a}$$

$$\frac{d}{dt}S(t) = \mu S^*(t) - (\beta\phi I(t) + \omega + \lambda)S(t) + \gamma R(t) \tag{4b}$$

$$\frac{d}{dt}S^*(t) = \lambda S(t) - (\omega + \mu)S^*(t) \tag{4c}$$

$$\frac{d}{dt}I(t) = \omega S(t) + (\beta\phi S(t) - \delta - \lambda)I(t) + \mu I^*(t) \tag{4d}$$

$$\frac{d}{dt}I^*(t) = \omega S^*(t) + \lambda I(t) - (\delta + \mu)I^*(t) \tag{4e}$$

$$\frac{d}{dt}R(t) = \delta I(t) - (\gamma + \lambda)R(t) + \mu R^*(t) \tag{4f}$$

$$\frac{d}{dt}R^*(t) = \delta I^*(t) + \lambda R(t) - \mu R^*(t) \tag{4g}$$

$$0 \leq S(t), S^*(t), I(t), I^*(t), R(t), R^*(t) \leq 1. \tag{4h}$$

5.1.1 *Stationary Regime Analysis*

We write system (4) succinctly as $\frac{d}{dt}y = F(y)$. We study the stationary regime of the system, i.e., the regime where, for $t \longrightarrow \infty$, the network state does not change with time. In particular, we look for equilibrium points, i.e., network states at which $\frac{d}{dt}y = 0$.

Setting $F(y) = 0$ and solving for y, we reach the following system of nonlinear equations.

$$S^* = iS \tag{5a}$$

$$I = \frac{aS}{bS + c} \tag{5b}$$

$$I^* = (\frac{gS + e}{bS + c})S \tag{5c}$$

$$R = (\frac{-dgS + f}{bS + c} + h)S + d \tag{5d}$$

$$R^* = 1 - (\frac{g(1-d)S + a + e + f}{bS + c} + i + h)S - d \tag{5e}$$

$$jS^2 + kS + cd\gamma = 0, \tag{5f}$$

where

$$a = -\omega\mu(\lambda + \mu + \omega + \delta) - \delta\omega^2 \tag{6a}$$
$$b = \beta\phi(\mu(\delta + \mu + 1) + \delta\omega) \tag{6b}$$
$$c = -\delta(\mu + \omega)(\delta + \lambda + \mu) \tag{6c}$$
$$d = \mu(\lambda + \mu + \gamma)^{-1} \tag{6d}$$
$$e = -\omega\lambda(\lambda + \mu + \omega + \delta) \tag{6e}$$
$$f = \omega(\lambda + \mu + \omega + \delta)(\lambda + \mu - \delta)d \tag{6f}$$
$$g = \omega\lambda\beta\phi \tag{6g}$$
$$h = -d(\lambda + \mu + \omega) \tag{6h}$$
$$i = \lambda(\omega + \mu)^{-1} \tag{6i}$$
$$j = (hb - dg)\gamma - a(\beta\phi) - b(\omega + \lambda - i\mu) \tag{6j}$$
$$k = (f + hc + bd)\gamma - c(\omega + \lambda - i\mu) \tag{6k}$$

Having expressed all variables in terms of S, we need to solve the quadratic equation (5f) for S, keeping in mind that any solution S_0 has to satisfy $0 \leq S_0 \leq 1$. The value of S_0 can be found from the quadratic formula:

$$S_0 = \frac{1}{2j}\left(-k \pm \sqrt{k^2 - 4jcd\gamma}\right) \tag{7}$$

Then, we substitute S_0 into (5a)-(5e) to find the other values $S_0^*, I_0, I_0^*, R_0, R_0^*$.

So, we found the only admissible equilibrium point of the network. We now give a sufficient condition for this point to be locally asymptotically stable, that is, all system trajectories starting near enough to the equilibrium point will eventually converge to it without wandering too far away in the meantime. This condition is that the Jacobian matrix of the system, evaluated at the equilibrium point, has eigenvalues with strictly negative real parts. Note that, instead of using the differential equation for R^*, we substitute $R^* = 1 - S - S^* - I - I^* - R$ and compute the Jacobian of an equivalent system with only the 5 variables S, S^*, I, I^*, R. The Jacobian $J(S, I)$ is

$$\begin{pmatrix} -\beta\phi I - \omega - \lambda & \mu & -\beta\phi S & 0 & \gamma \\ \lambda & -\omega - \mu & 0 & 0 & 0 \\ \beta\phi I + \omega & 0 & \beta\phi S - \delta - \lambda & \mu & 0 \\ 0 & \omega & \lambda & -\mu - \delta & 0 \\ -\mu & -\mu & \delta - \mu & -\mu & -\gamma - \lambda - \mu \end{pmatrix} \tag{8}$$

which, as we see, is only a function of S and I. The eigenvalues of $J(S, I)$ evaluated at the equilibrium point can be found by solving the 5th order equation

$$|J(S_0, I_0) - xI_5| = 0 \tag{9}$$

for x, where I_5 is the 5×5 unit matrix. As we have mentioned, if all the solutions have a strictly negative real part, then the equilibrium point is locally asymptotically stable. Moreover, if all the solutions have a strictly negative real part, the equilibrium point persists under small perturbations of the system parameters. That is, if $v(y)$ is any smooth vector field on $\mathbb{R}^5$, then for sufficiently small ϵ the equation

$$\frac{d}{dt}y = F(y) + \epsilon v(y) \tag{10}$$

has an equilibrium point near the original one, and the equilibrium point of the perturbed system is also locally asymptotically stable.

In Section 7, we show that all the eigenvalues have strictly negative real part for the range of system parameters we consider; hence, the equilibrium point is stable, and it persists under small perturbations of the system parameters. The stability analysis justifies using the equilibrium point to evaluate our system. If it were unstable, then either the system would not converge to it or the smallest disturbance would cause the system to leave it.

5.1.2 Time-dependent mobility

So far, we have assumed that user mobility, expressed through parameters μ, λ, and β, does not change with time. But mobility is usually time-dependent and periodic: users have different mobility pattern in the morning than in the afternoon, but these patterns repeat almost everyday. To address the time-dependence of mobility, we can split time into time periods and compute the mobility parameters for each time period separately.

Making μ, λ, and β time-dependent in (4) means that there is no longer an equilibrium point, because the fraction of users in each state (e.g., Seeker, Informed, Removed) continuously changes over time. We solve this system of nonlinear differential equations using numerical methods (as it is difficult to find their closed-form solutions), which provide us with the fraction of users at each time unit.

5.2 Baseline MobiCrowd: Buffer Only

To be able to isolate the effect of collaboration, we study the case where there is no collaboration among users and MobiCrowd relies only on its buffer to protect users' privacy: A user who becomes interested checks her buffer, and if the content is not there, she immediately contacts the server. Thus, there are no Seeker (S and S^*) users in the model for this case:

$$I(t) + I^*(t) + R(t) + R^*(t) = 1 \tag{11a}$$
$$\frac{d}{dt}I(t) = \gamma R(t) + \mu I^*(t) - (\lambda + \delta)I(t) \tag{11b}$$
$$\frac{d}{dt}I^*(t) = \lambda I(t) - (\mu + \delta)I^*(t) \tag{11c}$$
$$\frac{d}{dt}R(t) = \delta I(t) + \mu R^*(t) - (\lambda + \gamma)R(t) \tag{11d}$$
$$\frac{d}{dt}R^*(t) = \delta I^*(t) + \lambda R(t) - \mu R^*(t) \tag{11e}$$
$$0 \leq I(t), I^*(t), R(t), R^*(t) \leq 1. \tag{11f}$$

For the stationary regime analysis, we compute the equilibrium point of the system, and study its stability as before.

$$I^* = zI \tag{12a}$$

$$R = -\frac{1}{\gamma}(z\mu - \lambda - \delta)I \tag{12b}$$

$$R^* = 1 - I\left(1 - \frac{1}{\gamma}(z\mu - \lambda - \delta) + z\right) \tag{12c}$$

$$I = \mu\left(\frac{\delta}{\gamma}(\lambda + \gamma + \mu)(1 + z) - \delta + \mu(1 + z)\right)^{-1} \tag{12d}$$

where $z = \lambda(\mu + \delta)^{-1}$.

To compute the stability of this point, we compute the Jacobian for an equivalent system which arises after substituting $R^* = 1 - I - I^* - R$. In this case the system is linear, so if the eigenvalues are negative, then the equilibrium point is globally asymptotically stable, that is, the system converges to it for any initial condition. The Jacobian is

$$J = \begin{pmatrix} -\lambda - \delta & \mu & \gamma \\ \lambda & -\mu - \delta & 0 \\ \delta - \mu & -\mu & -\mu - \lambda - \gamma. \end{pmatrix} \tag{13}$$

The equation to solve for the eigenvalues is, similarly as before, $|J - xI_3| = 0$. We will show the stability of the equilibrium point in the next section.

For time-dependent mobility parameters, as before, we analyze the system numerically.

6 Quantitative Analysis

The direct objective of MobiCrowd is to hide user queries from the server. We quantify this objective, as our first evaluation metric, through the *hiding probability*: the probability that a user's query becomes hidden from the server due to MobiCrowd protocol. Under various user mobility and information spreading dynamics, we compute this metric using the results of the time-dependent epidemic model, and we compare to the results of simulations on a dataset of real mobility traces. In Section 7, we show that the simulation results corroborate our model-based findings about the hiding probability.

As our second evaluation metric, we quantify the *location privacy* that Mobicrowd offers to users against localization attacks. Specifically, we compute the expected error of an adversary who observes a user's trace and then forms a probabilistic estimate of her location. This probabilistic estimate is based on a Bayesian location inference approach [30] that enables us to incorporate both the background knowledge and observation of the adversary and to precisely quantify the location privacy of users. We link this Bayesian inference to our epidemic model, by computing the observation probability of the adversary from the hiding probability of MobiCrowd.

6.1 Probability of Hiding in the Mobile Crowd

The *hiding probability* in a given region is estimated as the fraction of queries per time unit that are *not* observed by the server. The higher this fraction, the lower the adversary's success in performing inference attacks on the observed queries. Hiding some of the users' locations from the adversary has two benefits: (1) Users become less traceable over space and time, as observed queries from a user are sparser, hence harder to correlate with each other and easier to be confused with the queries of other users [20], [37], [8]. (2) The set of a user's observed queries becomes harder to link to the user's real name. The hiding probability can show the reduction in the amount of information the adversary obtains from the users' queries *compared to* the case where users directly contact the server for each query.

In the case of no collaboration among users, i.e., in *buffer-only* MobiCrowd, the users can retrieve the information either from their buffer or from the server. Only the I users have the information in their buffers, whereas the R users are forced to contact the server when they become interested. The I users ask queries at a total rate of γI, and the R users at a total rate of γR. Therefore, the hiding probability in this case is

$$HP_0 = \frac{I}{I + R} \tag{14}$$

where I and R are computed from (11).

In the case of *collaboration* with probability $\phi > 0$ among users, queries can also be answered by peers. Only an insider user who is not already a Seeker, i.e., Insider Informed and Insider Removed users, can send a new query. So, we focus only on them and we compute the hiding probability as the probability that the user's query, given that she is an Insider Informed/Removed, is answered by buffer or a peer.

The user is Insider Informed with probability $\frac{I}{I+R}$. By definition, the query of an Insider Informed user is immediately answered by the buffer. So, her hiding probability is 1.

Turning to Insider Removed users, the probability of being Insider Removed is $\frac{R}{I+R}$. By definition, such a user (who, right after sending the query, becomes an Insider Seeker) needs to wait for an Insider Informed peer to collaborate with her. If she cannot find one before her waiting time expires, she has to expose her location to the server. Which of the two happens first can be modeled as a competition between two exponential random processes: P with mean $1/\beta\phi I$, representing the time to get the response from peers, and S with mean $1/\omega$, representing the time to get the response from the server. Then, the hiding probability is the probability that process P wins:

$$\Pr\{P < S\} = \int_{-\infty}^{\infty} f_S(s)ds \int_{S-P>0} f_P(p)dp = \\ = \frac{\beta\phi I}{\beta\phi I + \omega} \quad (15)$$

So, finally, we compute the hiding probability as

$$HP_\phi = \frac{I}{I+R} + \frac{R}{I+R} \cdot \frac{\beta\phi I}{\beta\phi I + \omega} \quad (16)$$

where I and R are computed from (4). We can see that if we set the collaboration probability ϕ to zero, the hiding probability becomes equal to (14).

6.2 Location Privacy versus Inference Attacks

In a *localization attack* the adversary targets a specific user at a specific time instant and computes the probability distribution over the regions where the user might be [38]. This distribution is computed given the observed traces from the user. Formally, the adversary computes $\Pr\{\mathbf{A}_u^t = r|\mathbf{o}_u\}$ for user u at time instant t for all regions r, where $\mathbf{A}_u^t$ is the random variable for the actual location of user u at time t, and $\mathbf{o}_u$ is the observed trace from user u. In the case of MobiCrowd users, the (server's) observation at a time t is either null or the true location of the user. From the adversary's localization probability distribution, we quantify the location privacy of a user as *the probability of error of the adversary in guessing the user's true location*, averaged over all times t.

We use Bayes' rule to compute the localization probability for the adversary.

$$\Pr\{\mathbf{A}_u^t = r|\mathbf{o}_u\} = \frac{\Pr\{\mathbf{A}_u^t = r, \mathbf{o}_u\}}{\Pr\{\mathbf{o}_u\}} \quad (17)$$

$$= \frac{\Pr\{\mathbf{A}_u^t = r, \mathbf{o}_u^{1:t}\} \Pr\{\mathbf{o}_u^{t+1:T}|\mathbf{A}_u^t = r\}}{\Pr\{\mathbf{o}_u\}}$$

where T is the length of the observed trace; note also that we have used the conditional independence of $\mathbf{o}_u^{t+1:T}$ and $\mathbf{o}_u^{1:t}$ given $\mathbf{A}_u^t = r$. The probabilities in the numerator can be computed recursively using the forward-backward algorithm of Hidden Markov Models (HMM). The normalizing factor $\Pr\{\mathbf{o}_u\}$ can also be computed simply by summing the numerator over all regions r [30].

To compute (17) we need, according to the theory of HMM, two quantities: (1) the transition probability between regions (i.e., $p_u(r_j|r_i)$, mobility model of the user), and (2) the observation probability (i.e., $\Pr\{\mathbf{O}_u^t|\mathbf{A}_u^t = r\}$, the probability of each possible observation, given the true location of the user). We compute the observation probability from the MobiCrowd hiding probability (16) as

$$\Pr\{\mathbf{O}_u^t = o|\mathbf{A}_u^t = r\} = \begin{cases} 1-\gamma(1-HP_\phi) & o = \text{null} \\ \gamma(1-HP_\phi) & o = r \\ 0 & o.w. \end{cases} \quad (18)$$

Having specified the transition and observation probabilities, we run the forward-backward algorithm (for hidden Markov models) to compute the localization probabilities for each time t. We then compute their average value over all time units t to compute the location privacy of users of our privacy-preserving scheme for various system parameters.

7 Evaluation

The location traces that we use belong to 509 randomly chosen mobile users (vehicles) from the epfl/mobility dataset at CRAWDAD [39]. We set the time unit of the simulation to 5 minutes and we consider the users' locations at integer multiples of the time unit, hence synchronizing all the traces. We group time units into three equal-size time periods: morning, afternoon, evening. We divide the Bay Area into 10×25 equal-size regions. Two users in a region are considered to be neighbors of each other if they are within 100m of each other (using WiFi). We run our simulation for 100 times on the traces and compute the average of the results.

From the location traces, we construct the time-dependent mobility model of each individual user, in the format of transition probability matrices (one matrix per time period). We also compute the average mobility model, which reflects how the whole crowd moves. For each region and time period we compute the mobility parameters λ, μ, and β separately (see Section 5).

We set the average waiting time before contacting the server, $1/\omega$, to 1, in effect choosing it as the unit by which the information lifetime and the request rate will be measured. We evaluate the system for all combinations of collaboration level $\phi = \{0.5, 1\}$, information lifetime $1/\delta = \{1, 4, 10, 16, 22\}$, and request rate $\gamma = \{0.05, 0.2, 0.4, 0.6, 0.8\}$. Information lifetimes lower than 1, i.e. shorter than the waiting time, do not make much sense. If information expires fast, the user cannot be willing to wait a long time before getting it, as it would be stale by the time it were received. Similarly, request rates larger than one imply multiple requests per time unit. But this cannot be compatible with the user's willingness to wait for one time unit.

7.1 Validation of Epidemic Model

In order to validate our model, we compare our numerical computation of hiding probability with simulation results.

The mobility parameters β, λ, μ and the ranges of system parameters γ, δ, ϕ are plugged into the epidemic model of MobiCrowd in order to compute numerically the solutions of (4) and (11) as functions of time. In other words, we compute the fraction of users in each state for each time unit. Note that this is different from just computing the stationary regime solutions. We then compute the hiding probability

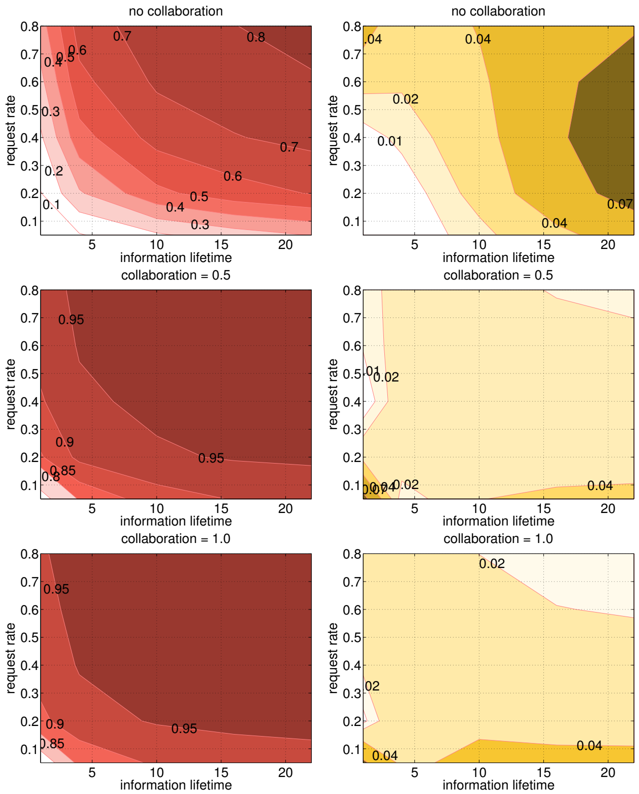


Fig. 1. Users' hiding probability, due to MobiCrowd, for the region under study (in downtown San Francisco). The first row illustrates the hiding probability of users when there is no collaboration, i.e. when users have to contact the server if they do not find the sought information in their buffer. The second and third rows show the same metric for collaboration factors $\phi = 0.5$ and $\phi = 1$, respectively. The left column shows the numerical results obtained from the epidemic model, whereas the right column shows the difference between the model and the simulation results.

as a function of time from (14) and (16). Finally, we simulate the MobiCrowd protocol on the location traces, and for each region and each time unit, we compute the users' hiding probability using directly its definition as the fraction of queries hidden from the LBS server. We plot the results obtained in Figure 1.

Figure 1 illustrates the average users' hiding probability using MobiCrowd with and without collaboration (PG_ϕ and PG_0). As it is not possible to plot the results for all the regions, we compute, as a representative example, the hiding probability in one region, located in downtown San Francisco. It has a higher concentration of points of interest, and 90 users are present in it on average, with a contact rate of $\beta = 51.89$ per user per time unit. The results of the numerical evaluation are displayed side by side with their absolute difference with the simulation results. This enable us to verify the validity of our epidemic model. The qualitative and also quantitative match between the simulation and the model enables us to rely on our epidemic model to evaluate users' location privacy in a very computationally efficient way in complex scenarios dealing with large networks.

All the plots confirm a general pattern of increasing hiding probability as the information lifetime or the request rate increases. With either kind of increase, users retrieve with higher probability non-expired information either from their own buffer or from their peers; hence, a higher fraction of their queries will be hidden from the LBS. Moreover, the hiding probability of each query for long lifetimes and low request rate values (i.e., long intervals between requests) appears to be more or less the same as the hiding probability for short lifetimes and high request rate values (i.e., short intervals between requests), as indicated by the vaulted shape of the contours. Also, adding collaboration to the buffering technique in MobiCrowd increases the fraction of hidden queries even for a collaboration factor of $\phi = 0.5$.

7.2 Evaluation of Privacy

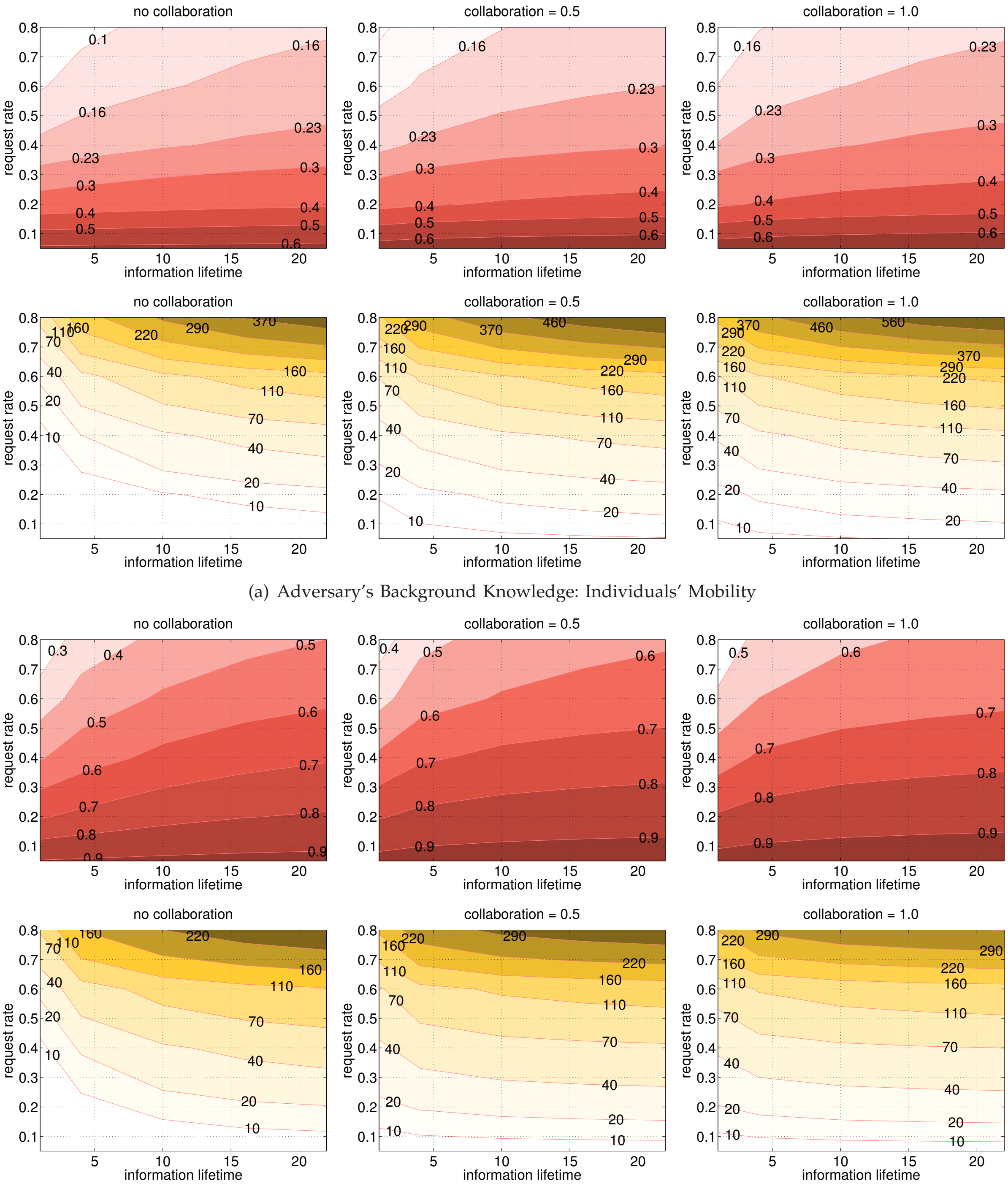
We use Location-Privacy Meter [8] to quantify the location privacy of users as the expected error of the adversary in guessing their correct location, including at times when they *do not* issue a query, i.e. between two successive LBS queries. We are interested in analyzing the privacy effect of the following factors:

- The adversary's background knowledge on user mobility, which can be
  - the mobility model of each individual user (**Individuals' Mobility Model**), or
  - the average mobility model of the whole user population (**Average Mobility Model**).
- The adversary's method of attack, which can consist of
  - just observing exposed locations, i.e. not try to guess a user's locations between two queries (**Observation adversary**), or
  - perpetrating Bayesian localization attacks to infer the whole location trace of each user (**Inference adversary**).

We compute privacy for multiple combinations of these factors, with and without our protection protocol. These are the concrete scenarios we study:

- **Baseline**: Inference without observations
- **No Protection** vs. Observation/Inference
- **MobiCrowd** vs. Observation/Inference

In the *Baseline* scenario, we compute privacy against the inference attack, assuming that the adversary ignores his LBS observations, relying only on his background knowledge. This scenario quantifies the extent to which the adversary's knowledge is by itself sufficient to predict the users' locations over time. It is a baseline scenario, in the sense that no privacy mechanism can achieve better privacy than this.



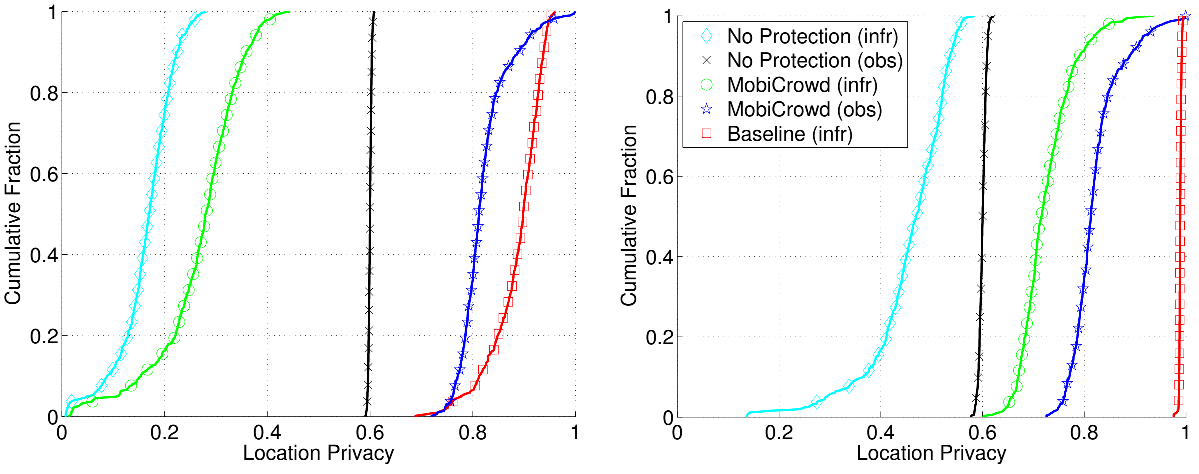
(a) Adversary's Background Knowledge: Individuals' Mobility

(b) Adversary's Background Knowledge: Average Mobility

Fig. 2. Average Location Privacy of MobiCrowd users against the Bayesian inference localization attack (top row of each sub-figure), and the %-improvement that MobiCrowd achieves over no protection, when MobiCrowd is not in place (bottom row of each sub-figure). The considered adversary's background knowledge is the set of mobility models of all individual users, in Sub-Fig. (a), and the average mobility model of all users, in Sub-Fig. (b).

In the *No Protection* scenario, users submit their queries directly and immediately to the server without using any protection mechanism. This scenario reflects the risk of unprotected use of LBSs. We compute privacy against the observation and against the inference adversaries.

In the *MobiCrowd* scenarios, we again compute privacy against the observation/inference adversaries. However, in this case, users make use of MobiCrowd, hence their observed traces contain fewer number of locations than in the no protection scenario.



(a) Background Knowledge: Individuals' Mobility

(b) Background Knowledge: Average Mobility

Fig. 3. Cumulative Fraction of users' location privacy in different protection/attack scenarios. Users' collaboration level is 0.5, the request rate is 0.4, and the information lifetime is 10. The graphs show what fraction of users (on the y-axis) have a privacy level up to a certain point (on the x-axis). Sub-figures (a) and (b) differ in terms of the background knowledge of the adversary (used in the Bayesian inference attack). The *Baseline (infr)* graph shows their location privacy against the Bayesian inference attack, if the adversary relies only on his background knowledge. The *No Protection (infr)* graph shows users' location privacy against the Bayesian inference attack, if they do not use any protection mechanism and submit their queries to the server. The *No Protection (obs)* graph shows location privacy of users in terms of the fraction of times their true location is not exposed to the server, because they didn't have any query. The *MobiCrowd (infr)* shows location privacy of MobiCrowd users against the Bayesian inference attack. The *MobiCrowd (obs)* shows location privacy of MobiCrowd users in terms of the fraction of times their true location is not exposed to the server, due to the protection or lack of a query.

7.2.1 Average Location Privacy

To see how our system performs across a range of parameters, we compute, for all combinations of system parameters (request rate γ, information lifetime $1/\delta$, and collaboration probability ϕ), the average location privacy of users against the localization attack, as explained in Sec. 6.2, for the MobiCrowd and No Protection scenarios.

Fig. 2 shows the location privacy of MobiCrowd users against the localization attack, as well as the %-improvement of their privacy over having no protection (i.e. when they send all their queries to the server). Fig. 2(a) and Fig. 2(b) illustrate the results for the cases where the adversary's knowledge is the mobility model of all individual users, and their average mobility model, respectively. Thus, the comparison between Fig. 2(a) and Fig. 2(b) shows the effect of the adversary's background knowledge on the users' location privacy.

MobiCrowd achieves the best %-improvement in the high (> 0.6) request rate regime, especially if the information lifetime is not too low. If the request rate is low, few locations are exposed in the first place, so location privacy is already high even without protection. Privacy is in danger at high request rates, where MobiCrowd's improvement is significant: It ranges from 2x (100%) up to 6.5x (550%). This observation holds true across all twelve cases in Fig. 2.

As expected, the adversary does considerably better when using each user's own mobility model in the attack, rather than using the average mobility model for everyone. More precisely, the success probability of our Bayesian inference attack, in estimating a user's location between two successive observations, significantly increases if we provide the adversary with a more precise mobility model. However, we see that MobiCrowd here again helps when it is most needed, and significantly improves (up to 550%) the users' location privacy when the adversary is very powerful due to his accurate background knowledge.

Finally, note that, although more collaboration is definitely better, full collaboration $\phi = 1$ is not necessary to reap the benefits of MobiCrowd. Even at $\phi = 0.5$ there is a considerable privacy gain.

The only cases where MobiCrowd's improvement is below 100% is when privacy is already high, in which case a further increase does not really matter, or when information expires too fast, in which case the users are forced to contact the server for most of their queries.

7.2.2 Cumulative Distribution of Location Privacy

In order to better analyze the added value of the adversary's knowledge and his inference attack on the one hand, and the effectiveness of MobiCrowd on the other hand, we compute users' location privacy for all the scenarios we enumerated in Sec. 7.2, but for a single set of parameters ($\gamma = 0.4$, $\delta = 0.1$, and $\phi = 0.5$). We plot the results in Figure 3, which shows the *cumulative distributions* of users' location privacy in different scenarios. Plotting cumulative distributions allows us to observe Mobicrowd's improvements for all desirable percentiles of users, instead of being limited to the previously computed averages over all users.

The baseline privacy in Fig. 3(a) and Fig. 3(b) show how much information is embedded in the background knowledge of the adversary, i.e., how accurately he can predict users' location, relying only on their mobility model.

In each of the sub-figures, the Baseline (inference) and No Protection (inference) scenarios reflect the *risk* of using location-based services without any protection. Even an adversary with knowledge of the average mobility can significantly decrease users' location privacy, hence the extreme need to employ privacy enhancing protocols such as MobiCrowd.

The difference, approximately 35%, between location privacy in MobiCrowd (observation) and No Protection (observation) shows the added value of MobiCrowd with respect to an observer (e.g., a curious but not adversarial LBS operator). However, these privacy values do not constitute a lower bound on user privacy, as an inference adversary can estimate the actual location of users more accurately.

We can see the additional damage caused by an inference adversary, compared to an observer, by comparing corresponding (observation) and (inference) scenarios. There is a difference of about 3x for the Individuals' Mobility Model, and a much smaller one, 15-30%, for the Average Mobility Model. This is to be expected, as the quality of the inference depends a lot on the quality of the background knowledge.

The added value of MobiCrowd against an inference adversary is about 50%, when the adversary's knowledge is Individual Mobility Model, and a bit less than 50% when the knowledge is Average.

7.3 Implementation

We implement MobiCrowd on three different Nokia mobile devices (N800, N810, and N900) by building a *mobile privacy proxy* in each device. The proxy does not require any modification of the supported applications and it is transparent to their operation. The prototype works with the Maemo Mapper LBS and MobiCrowd acts as an HTTP transparent proxy to which the client traffic is redirected. Note that knowing the format of the LBS queries and the data format of the server replies is enough to adapt MobiCrowd to new LBS applications. Our implementation in Python is 600 lines of code, including the proxy module, ad-hoc networking module, and the server interface module. Memory utilization does not exceed 3% of the total device memory.

We perform measurements on a 5-device testbed to estimate the delay to obtain a peer response. Three out of the five are randomly chosen to collaborate each time. Mobiles access the LBS server over a cellular link (e.g., GSM), and communicate with each other via the WiFi interface. Averaged over 100 queries, the delay is 0.17sec. We also note that cryptographic delays are (for a typical OpenSSL distribution) low: the weakest of the three devices, the N800, can verify more than 460 RSA signatures per second (1024 bit), or 130 signature verification per second (for 2048 bit modulus); this implies that digitally signed LBS response can be easily handled by the devices to protect against malicious peers.

A popular technique that enhances privacy against local eavesdroppers is to change the identifiers frequently. Cellular network operators make use of *network-issued pseudonyms* (TMSIs) to protect the location-privacy of their users [32]. MobiCrowd-ready mobile devices can also mimic this defense (as has already been proposed for wireless networks, e.g., [31]). They can change their identifiers (e.g., the MAC addresses) as often as desired, even while in a single point-of-interest area. This would essentially root out any threat by any curious local observer. Even in the case of a stalker, it would not be possible to link the successive identifiers of a device to that device, as multiple users' identifiers will be mixed together. The only remaining option for the stalker is to maintain visual contact with the target user, but defending against this threat is clearly orthogonal to our problem.

8 Conclusion

We propose a novel approach to enhance the privacy of LBS users, aiming against service providers who could extract information from their LBS queries and misuse it. We develop and evaluate MobiCrowd, a scheme that allows LBS users to hide in the crowd and to reduce their exposure while they continue to receive the location context information they need. MobiCrowd achieves this by leveraging the collaboration between users, who have the incentive and the capability to safeguard their privacy. We propose a novel analytical framework to quantify location privacy of our distributed protocol. Our epidemic model captures the hiding probability for user locations, i.e. the fraction of times when, due to MobiCrowd, the adversary does not observe user queries. By relying on this model, our Bayesian inference attack estimates the location of users when they hide. Our extensive joint epidemic/Bayesian analysis shows a significant improvement thanks to MobiCrowd, across both the Individual and the Average Mobility background knowledge scenarios for the adversary. We demonstrate the resource efficiency of MobiCrowd by implementing it in portable devices.

References

[1] "Pleaserobme: http://www.pleaserobme.com."

[2] J. Meyerowitz and R. Roy Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *MobiCom '09: Proceedings of the 15th annual international conference on Mobile computing and networking*, 2009.

[3] F. Olumofin, P. K. Tysowski, I. Goldberg, and U. Hengartner, "Achieving efficient query privacy for location based services," in *Privacy Enhancement Technologies (PETS)*, 2010.

[4] G. Ghinita, P. Kalnis, A. Khoshgozaran, C. Shahabi, and K.-L. Tan, "Private queries in location based services: anonymizers are not necessary," in *Proceedings of the ACM SIGMOD international conference on Management of data*, 2008.

[5] R. Anderson and T. Moore, "Information Security Economics–and Beyond," *Advances in Cryptology-CRYPTO*, 2007.

[6] R. Shokri, J. Freudiger, M. Jadliwala, and J.-P. Hubaux, "A distortion-based metric for location privacy," in *WPES '09: Proceedings of the 8th ACM workshop on Privacy in the electronic society*. New York, NY, USA: ACM, 2009, pp. 21–30.

[7] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "A parsimonious model of mobile partitioned networks with clustering," in *Proceedings of the First international conference on COMmunication Systems And NETworks*, 2009.

[8] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *IEEE Symposium on Security and Privacy*, Oakland, CA, USA, 2011.

[9] J. Krumm, "A survey of computational location privacy," *Personal Ubiquitous Comput.*, vol. 13, no. 6, pp. 391–399, 2009.

[10] R. Shokri, J. Freudiger, and J.-P. Hubaux, "A unified framework for location privacy," in *HotPETs*, 2010.

[11] R. Shokri, P. Papadimitratos, G. Theodorakopoulos, and J.-P. Hubaux, "Collaborative location privacy," in *8th IEEE International Conference on Mobile Ad-hoc and Sensor Systems (IEEE MASS 2011)*, October 2011.

[12] R. Shokri, P. Papadimitratos, and J.-P. Hubaux, "Mobicrowd: A collaborative location privacy preserving LBS mobile proxy (demonstration)," in *8th ACM international conference on Mobile systems, applications, and services (MobiSys) - Demo Session*, 2010.

[13] "NIC: Nokia Instant Community."

[14] "Wi-Fi Direct: http://www.wi-fi.org/wi-fi_direct.php."

[15] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "GreenGPS: a participatory sensing fuel-efficient maps application," in *ACM MobiSys*, 2010.

[16] Y. Liu, A. Rahmati, Y. Huang, H. Jang, L. Zhong, Y. Zhang, and S. Zhang, "xShare: supporting impromptu sharing of mobile phones," in *Proceedings of the 7th international conference on Mobile systems, applications, and services*, 2009.

[17] J. Freudiger, R. Shokri, and J.-P. Hubaux, "Evaluating the privacy risk of location-based services," in *Financial Cryptography and Data Security*. Springer, 2012, pp. 31–46.

[18] M. Srivatsa and M. Hicks, "Deanonymizing mobility traces: Using social network as a side-channel," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 628–637.

[19] N. Vratonjic, K. Huguenin, V. Bindschaedler, and J.-P. Hubaux, "How others compromise your location privacy: The case of shared public IPs at hotspots," in *13th Privacy Enhancing Technologies Symposium (PETS)*, 2013.

[20] B. Hoh and M. Gruteser, "Protecting location privacy through path confusion," in *Proceedings of the First International Conference on Security and Privacy for Emerging Areas in Communications Networks*, 2005.

[21] P. Golle and K. Partridge, "On the anonymity of home/work location pairs," in *Proceedings of the 7th International Conference on Pervasive Computing*, 2009.

[22] A. R. Beresford and F. Stajano, "Mix zones: User privacy in location-aware services," in *PERCOMW '04: Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops*. Washington, DC, USA: IEEE Computer Society, 2004, p. 127.

[23] J. Freudiger, R. Shokri, and J.-P. Hubaux, "On the optimal placement of mix zones," in *PETS '09: Proceedings of the 9th International Symposium on Privacy Enhancing Technologies*. Berlin, Heidelberg: Springer-Verlag, 2009, pp. 216–234.

[24] C.-Y. Chow, M. F. Mokbel, and X. Liu, "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," in *GIS '06: Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*. New York, NY, USA: ACM, 2006, pp. 171–178.

[25] R. Shokri, C. Troncoso, C. Diaz, J. Freudiger, and J.-P. Hubaux, "Unraveling an old cloak: k-anonymity for location privacy," in *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society*, 2010.

[26] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *ACM CCS*, 2013.

[27] R. Chow and P. Golle, "Faking contextual data for fun, profit, and privacy," in *WPES '09: Proceedings of the 8th ACM workshop on Privacy in the electronic society*. New York, NY, USA: ACM, 2009, pp. 105–108.

[28] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *ACM CCS*, 2012.

[29] F. Santos, M. Humbert, R. Shokri, and J.-P. Hubaux, "Collaborative location privacy with rational users," in *Decision and Game Theory for Security*. Springer, 2011, pp. 163–181.

[30] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, "Quantifying location privacy: the case of sporadic location exposure," in *Proceedings of the 11th international conference on Privacy enhancing technologies*, ser. PETS'11. Berlin, Heidelberg: Springer-Verlag, 2011.

[31] T. Jiang, H. J. Wang, and Y.-C. Hu, "Preserving location privacy in wireless LANs," in *MobiSys '07: Proceedings of the 5th international conference on Mobile systems, applications and services*. New York, NY, USA: ACM, 2007, pp. 246–257.

[32] 3rd Generation Partnership Project, "3GPP GSM R99," in *Technical Specification Group Services and System Aspects*, 1999.

[33] T. G. Kurtz, *Approximation of population processes*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1981.

[34] W. O. Kermack and A. G. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc R Soc Lond A*, vol. 115, pp. 700–721, 1927.

[35] G. Theodorakopoulos, J.-Y. Le Boudec, and J. S. Baras, "Selfish response to epidemic propagation," *IEEE Transactions on Automatic Control*, vol. 58, no. 2, pp. 363–376, February 2013.

[36] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Comput. Netw.*, vol. 51, pp. 2867–2891, July 2007. [Online]. Available: http://portal.acm.org/citation.cfm?id=1242848.1243153

[37] J. Krumm, "Inference attacks on location tracks," in *Pervasive '07: Proceedings of the 5th International Conference on Pervasive Computing, volume 4480 of LNCS*, 2007.

[38] R. Shokri, "Quantifying and protecting location privacy," Ph.D. dissertation, EPFL, 2013.

[39] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD data set epfl/mobility (v. 2009-02-24)."

Reza Shokri received his M.S. degree in computer engineering from University of Tehran, Iran, in 2007, and his Ph.D degree in communication science from EPFL, Switzerland, in 2013. His research focuses on quantitative privacy.

George Theodorakopoulos received the Diploma degree from the National Technical University of Athens, Greece, in 2002, and the M.S. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 2004 and 2007, all in electrical and computer engineering. His research interests are in privacy, security and trust in networks.

Panos Papadimitratos is an Associate Professor at KTH, Stockholm, in the School of Electrical Engineering and its Communication Networks division.

Ehsan Kazemi is a PhD student at LCA4, EPFL. His research is focused on complex networks, data analysis, and privacy. He received both his B.S. and M.S. degrees in Communication Systems from Sharif University of technology, IRAN.

Jean-Pierre Hubaux is a professor in the School of Computer and Communication Sciences of EPFL and has pioneered research areas such as the security of mobile ad hoc networks and of vehicular networks. He is currently working on data protection in mobile communication systems and healthcare systems. He is a Fellow of both ACM and IEEE.