Kassem Fawaz*, Kyu-Han Kim, and Kang G. Shin

Privacy vs. Reward in Indoor Location-Based Services

Abstract: With the advance of indoor localization technology, indoor location-based services (ILBS) are gaining popularity. They, however, accompany privacy concerns. ILBS providers track the users' mobility to learn more about their behavior, and then provide them with improved and personalized services. Our survey of 200 individuals highlighted their concerns about this tracking for potential leakage of their personal/private traits, but also showed their willingness to accept reduced tracking for improved service. In this paper, we propose PR-LBS (Privacy vs. Reward for Location-Based Service), a system that addresses these seemingly conflicting requirements by balancing the users' privacy concerns and the benefits of sharing location information in indoor location tracking environments. PR-LBS relies on a novel location-privacy criterion to quantify the privacy risks pertaining to sharing indoor location information. It also employs a repeated play model to ensure that the received service is proportionate to the privacy risk. We implement and evaluate PR-LBS extensively with various real-world user mobility traces. Results show that PR-LBS has low overhead, protects the users' privacy, and makes a good tradeoff between the quality of service for the users and the utility of shared location data for service providers.

Keywords: Location Privacy, Indoor localization, Differential Privacy

DOI 10.1515/popets-2016-0031
Received 2016-02-29; revised 2016-06-02; accepted 2016-06-02.

1 Introduction

Localization technologies [1], tailored for indoor spaces such as retail stores, malls, airports, museums, and hospitals, are gaining popularity. An *indoor service provider* (SP) (e.g., retail store owner) utilizes the customers' indoor location information to study their behaviors and infer their preferences and interests. In the best case, this should be a win-win for both customers and SPs; the SPs collect location data and deliver better service to the customers which leads to enhanced customer satisfaction and eventually increased revenues.

Unfortunately, indoor localization has not been realized to its full potential. Customer resistance is forcing SPs to either sideline the technology (e.g., Nordstrom ceased customer tracking after public outrage [2]) or rely solely on anonymous data collection [3]. Analyzing customers' location data anonymously prevents the service provider from offering them personalized services that would result in revenue-generation/increase .

To gain a better understanding of the users' perspectives towards indoor localization, we surveyed 200 shoppers in two major retailers: Walmart and Nordstrom. The survey shows that customers have both privacy and utility concerns.

A. Privacy Concerns: Users cited privacy concerns for not accepting this technology (consistent with other surveys [4]). An SP, tracking users' mobility, has the potential to infer personality traits and/or habits that could be private to them. For example, a retailer can infer from the frequently-visited aisles the shopper's gender (men's vs. women's clothing), ethnicity (ethnic food aisles), socioeconomic status (expensive vs. inexpensive clothing and accessories), health condition (pharmacy aisles), sensitive interests (sporting goods, adult magazines and films), religious beliefs (clothing, specific food aisles), etc.

Unlike the outdoor case, the indoor SP is directly involved in the user's localization through the deployed infrastructure such as Wi-Fi and Bluetooth. Unless the users turn off their devices, the SP does not provide an opt-out mechanism by which users can exert control over how much of their mobility is being tracked. According to our survey, users are not comfortable with the SP storing their mobility information even when it is processed anonymously.

B. Utility Concerns: Users expressed interest in receiving rewards for sharing some of their mobility information. This is referred to in the literature as a *fair transaction* [5]; a user shares some data proportionately

***Corresponding Author: Kassem Fawaz:** University of Michigan, E-mail: kmfawaz@umich.edu
Kyu-Han Kim: Hewlett Packard Labs, E-mail: kyu-han.kim@hpe.com
Kang G. Shin: University of Michigan, E-mail: kgshin@umich.edu

with the received rewards. In the indoor case, the users might find it challenging to engage in a fair transaction with the SP. First, it is hard for them to associate their mobility with a privacy cost in the typically public indoor space (in the outdoor case there is some notion of a private location such as home). Second, although the location information might help the user indirectly, through improving store layout, product placement, or waiting time in checkout lines, but these are not personalized and tangible services that would make users feel satisfied for revealing their mobility.

In this paper, we first pose a question: *can the seemingly conflicting requirements of the users and SPs be effectively resolved?* To answer this question, we propose PR-LBS (Privacy vs. Reward in Location Based Service); a novel framework that addresses the user's privacy concerns and enables them to receive the right reward from their location sharing on one side. On the other side, it provides the indoor SPs with enough information to perform aggregate and more personalized analysis of the customers.

PR-LBS puts the users in control, allowing them to specify a privacy setting that translates into a provable privacy guarantee. PR-LBS packs in an online private location release mechanism that achieves differential privacy guarantees in indoor environments. Additionally, PR-LBS enables the users to set high-level policies that provide their utility definition as a function of the privacy "cost" of sharing location and "benefit" received from the SP. To estimate the cost of sharing the user's indoor mobility, we introduce a new privacy criterion, which is based on information disclosure.

PR-LBS improves on the current approaches of take-it or leave-it; it ensures a fair transaction of the user's location information with the SP by abstracting the interactions between the user and the SP as a repeated play model [6]. PR-LBS employs the strategic experts algorithm [7] to choose, at run-time, the action (hide, reveal, or anonymize location) sequence that maximizes the user's utility.

PR-LBS is a generic framework that supports various practical deployment scenarios. A user can simply download and install it to the device, which we call *device mode*, if localization is device-based, such as iBeacons [8]. Also, a localization provider can employ PR-LBS as a broker between the user and the SP, which we call *infrastructure mode*, in case localization is infrastructure-based such as CUPID [1]. PR-LBS could act as a privacy guarantee/seal in this case [9], which will make users more comfortable to share their location. In this paper, we design and evaluate PR-LBS in both modes and present a full real-life implementation on Android for the device mode.

PR-LBS has a low energy footprint when running on the user's device and is easy to use as our user study (100 respondents) shows. Our survey also indicates that users are more comfortable with location tracking technology with PR-LBS being deployed. Further, our evaluations of PR-LBS in 8 different scenarios show that PR-LBS strikes a balance between the user and the SP. It controls the release of location information to protect the users' privacy, rewards them with commensurate service, and maintains data utility for the SP.

The paper is organized as follows. Section 2 reviews the related work. Section 3 presents our survey. Sections 4 and 5 present our system and privacy models, respectively. Section 6 details the design of PR-LBS . Section 7 describes our implementation and evaluation of PR-LBS. Section 8 lists some limitations of PR-LBS. Finally, Section 9 concludes the paper.

2 Related Work

There have been numerous efforts to mitigate the privacy risks in indoor environments. Retailers provide customers with opt-out options and claim to analyze their data in aggregate [3]. Our survey showed that users are likely to opt-out if provided with the option. Also, aggregate processing does little to protect the users' privacy. The SP still stores mobility data that is tagged with a MAC address (or a hashed form thereof). The hashed MAC can link the user to his traces [10] and can be reverse-mapped to the original MAC [11]. Alternatively, PR-LBS provides provable privacy guarantees by limiting the additional knowledge the SP attains from observing the user's mobility.

Other approaches rely on complete prevention of localization [12–15]. PR-LBS capitalizes on these mechanisms by acting as a control knob to opportunistically decide when to activate/deactivate them. PR-LBS exercises fine-grained location control to protect users' privacy while allowing them to interact with the SP. Recently, there have been mobile apps (such as that by Placed and Shopkick) that allow the users to receive rewards for location check-ins. PR-LBS automates this process; it acts on the user's behalf to decide when it is beneficial to share location or not. The user only specifies a privacy setting and high-level policy while PR-LBS takes care of deciding the privacy cost, service level, and sharing/hiding/anonymizing the location.

Researchers have proposed mechanisms to appraise private data before sharing it so that the user is properly rewarded (e.g., the architectures of Riederer *et al.* [16] and Ghosh and Roth [17]). These mechanisms typically require cooperation from both users and SPs. PR-LBS does not change the communication between the user and the SP and targets the case of a non-cooperative SP. Finally, Shokri [18] proposes a theoretical framework that optimizes user-side obfuscation to maintain both differential and distortion privacy with the impact on utility not being greater than the case of optimizing for a single privacy criterion. PR-LBS takes a different approach; it provides concrete mechanisms that achieve differential privacy. Moreover, it maximizes the user's rewards by adapting actions to the SP's services. PR-LBS is also a practical system that users can run in real-world environments.

3 Survey

We designed two surveys to study individuals' behavior and privacy preferences when shopping in Walmart (104 respondents) and Nordstrom (100 respondents) using Amazon Mechanical Turk. We compensated each participant with $3 and the average time for survey completion was 24 minutes. To protect the privacy of the participants, we did not collect any personal information and processed the data in aggregate. We also introduced a set of questions to weed out inconsistent responses.

The participants are diverse; they are uniformly distributed among genders, 42% are between 15 and 29 years old, while 43% fall between 30 and 44 years of age, half with a university/college degree, 29% with a high school degree, and 75% visit a brick-and-mortar shop at least once a week (97% at least once a month).

In the first section of the survey, we presented the respondents with the disclaimer (Fig. 1) that Nordstrom displayed to the shoppers in 2013. We then asked them, after reading the disclaimer, whether they would hypothetically consent to either Nordstrom or Walmart gathering Wi-Fi information assisting in their localization. This was the first mentioning of indoor localization-related terminology in the survey; we used the same language of a retailer to avoid any bias.

Interestingly, the participants responded with a preference to prevent the store from gathering information assisting in their localization (70%), 18% indicated that they would consent to the store gathering parts

> **We are always looking for ways to improve our customers' experience. We gather publicly-broadcasted information your smartphone or other WiFi-enabled device sends out when it is attempting to connect to a WiFi network in and around this store. This provides us with anonymous, aggregate reports that give us a better sense of customer foot traffic. We do not gather such things as your name, email address, phone number, your device's browsing activity or text, email or voice messages.**

Fig. 1. The disclaimer presented at the start of the survey.

of their Wi-Fi information. Only 10% of the participants consented to full gathering of Wi-Fi information broadcasted by their devices. We then posed the same question differently by indicating that the disclaimer effectively asks for the user's consent for indoor location tracking. The response distribution shifted a bit; 61% chose to prevent location tracking entirely, 24% chose to allow the store to gather part of their mobility, while the rest (15%) consented to full location tracking.

In the rest of this paper, we refer to the first set of individuals (rejecting tracking) as *privacy-oriented*, the second set (consenting to only part of tracking) as *neutral*, while we refer to the third set of individuals as *service-oriented*. These categories are akin to Westin's [19] categorization of privacy orientations of individuals as fundamentalists (privacy-oriented), pragmatists (neutral), and unconcerned (service-oriented).

Privacy-oriented Participants: These participants cited a set of privacy-related reasons as to why they reject location tracking. The most recurring reason was that they do not trust the store with mobility data (49%), the second being they do not feel comfortable with their mobility information being gathered (43%), and the third was that the store provides nothing in return for gathered mobility data (41%).

We also asked these respondents about their perception of the difference between smartphone-based and other tracking technologies such as monitoring purchase history (through credit card or rewards program) and using CCTV cameras. Regarding purchase history, only 10% indicated that purchase history reveals the same information as their mobility. The rest of the participants indicated that they do not want the store to know what items they are interested in but did not end up buying or that they use cash for their purchases. We observed a similar trend with CCTV cameras; only 20% of the participants indicated that cameras and location tracking reveal the same information while the others felt that it is harder to track them using CCTV cameras.

Neutral Participants: The second set of participants cited similar reasons as to why they want some part of

their mobility to be hidden. As for which parts of their mobility they want to be hidden, they responded with those areas that they deem private (65%), areas of the store that include items they browse but do not buy (44%), and areas where they receive nothing in return from the service provider (25%).

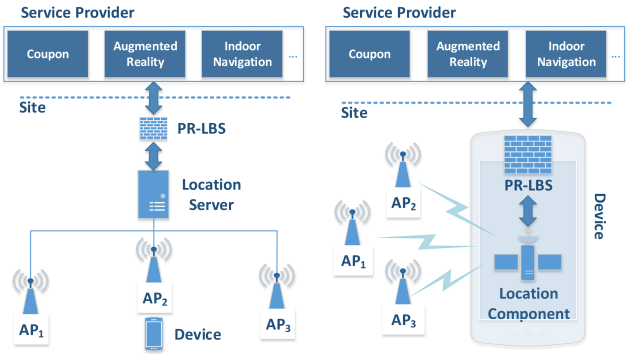
Service-oriented Participants: Individuals belonging to this set of survey respondents do not feel threatened by the store owner tracking their mobility. When asked whether they would change this perspective if the store owner would treat them differently based on mobility data, 30% indicated that would choose to prevent tracking, 37% still consented to full tracking, and the rest (33%) answered by not being sure. The participants' perspective further shifted when we indicated that the store owner might share their mobility with a third-party entity (an advertisement agency for example). 50% indicated that they do not consent to location tracking anymore and only 26% responded that they have no problem with their location being tracked.

In the second part of the survey, we asked participants to trace their path, on a map of the store, the last time they went shopping at either Walmart or Nordstrom if they remember it well. We then asked them to indicate the parts of the path they would hide from either store. Interestingly, even privacy-oriented respondents did not choose to hide all zones of their paths. In the last part of the survey, we asked participants to input their satisfaction level in different situations. More than 40% of the privacy-oriented users indicated that they would be satisfied if they were to share some of their mobility and receive very good service in return.

4 System Model

PR-LBS addresses the case of a user's location tracking **exclusively** in constrained public (including indoor) spaces, such as retail stores, malls, museums, theme parks, etc., where a localization system is installed. We consider the following main entities involved in the ecosystem:

- **User**: the individual moving around in the space of interest while carrying a mobile device.
- **Service Provider (SP)**: the entity owning the space in which the user moves. It manages a set of application servers that analyze the user's location and push service in the form of coupons, directions, deals, promos, etc.



(a) Infrastructure mode **(b)** Device mode

Fig. 2. PR-LBS deployment options

- **Localization Provider (LP)**: an entity contracted by the SP to localize the users. It relies on the deployed Wi-Fi access points or Bluetooth beacons to track users via the devices they carry. The LP can reside either on the device side or on the infrastructure side.

The SP deploys a mobile app (e.g., Shopkick) that acts as its communication channel with the user. The SP uses a consistent identifier such as the MAC address to map the location updates to the user running the app. The SP then pushes the tailored location-based content to the user through the app.

Logically, PR-LBS runs between the LP and the SP. It is a trusted module that controls the release of the location information to the SP in a privacy-aware manner. It is very important to note that the SP only views that mobility of the user that has been released by PR-LBS. PR-LBS runs in *device* or *infrastructure* mode:

Infrastructure mode (Fig. 2a): fits infrastructure-based localization where the LP has to install and run PR-LBS as the device can not control location sharing. This, however, could only happen if the SP has enough incentives to do so. Given users' privacy concerns, the SP has an incentive to deploy a solution that mitigates these concerns. For example, European companies have to apply for a privacy certification before collecting users' data (including indoor location) [9].

Device mode (Fig. 2b): fits device-based localization that computes the location on the device and then shares it with the SP. The user installs and runs PR-LBS that controls location release from the device.

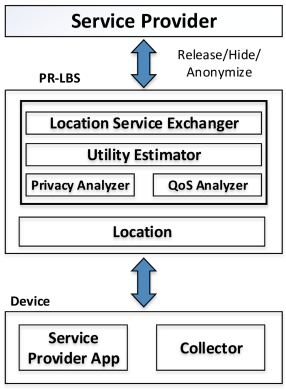


Fig. 3. The high-level operations of PR-LBS.

4.1 High-Level Description

Fig. 3 shows the high-level operations that PR-LBS performs when deployed in infrastructure mode. In device-mode deployment, PR-LBS performs the same operations while running on the device.

In the view of PR-LBS, the area of interest is partitioned into a set of zones: $Z = \{z_k\}$. The zones are semantic sub-areas within the area of interest which the SPs are typically interested in mapping the user's location to. For example, a zone could refer to an aisle in a supermarket, a department in a store, or an entertainment station in a theme park.

As the user moves from a zone z_i to another zone z_j, PR-LBS decides whether to *release*, *hide*, or *anonymize* z_i from the SP (Section 6.1). This action will result in a potential privacy cost to the user, $leak(z_i)$, as estimated by the *privacy analyzer* as described in Section 5.3. While spending time at z_i and then moving to z_j, the SP will be pushing a service to the user's device. When the user visits z_j, the *QoS analyzer* (Section 6.3) estimates the value of the service, $serv_i$, the user received as a result of hiding, releasing, or anonymizing z_i to the SP.

A *transaction* between the user and the SP takes place during the time period spanning the user reaching z_i and just before arriving at z_j. The *utility estimator* (Section 6.2) module of PR-LBS computes the utility (user-defined) the user gained after incurring a cost=$leak(z_i)$ and receiving a benefit= $serv_i$. The exchange module employs a set of privacy preserving mechanisms (Section 5.2) as "experts" that dictate the action to perform. This module utilizes the history of the user–SP transactions to decide the best expert (maximizing users' utility) to follow when reaching z_j.

Finally, PR-LBS has a *collector* module that runs on the device and collects the privacy preferences of the user. While PR-LBS is running, the collector module gathers information to assist the QoS analyzer in computing the service that the user receives.

5 Privacy Model

From the users' perspective, any entity collecting their location has the potential of posing privacy threats. We make a natural choice to trust the user's device as no solution is feasible without such a trust. We also choose to trust the LP only if it deploys PR-LBS, as it will indicate a willingness to provide privacy protection to the user. Thus, we assume that the privacy threats originate from the SP's analysis of the collected location data and the resulting treatment of the user.

The SP is an honest-but-curious entity that passively profiles the user through location information. These SPs will not collude with the LP (infrastructure-based case) if they choose to deploy PR-LBS. While the privacy threats are evident for an infrastructure-based LP, device-based localization might be perceived as less threatening. Proximity beacons can not track the user's mobility, but smartphone apps scanning for beacons pose tracking threats. We found that several shopping apps, including Shopkick, scan for nearby beacons and upload them along with consistent identifiers allowing them to track the user's mobility. In this paper, we only consider the threats to the user's privacy from location tracking that originate from the user's smartphone. An SP might utilize other channels to localize a user, e.g., CCTV cameras, which the user and PR-LBS, unfortunately, can not control. We also view security challenges as orthogonal to this work.

Intuitively, any privacy loss that the user suffers from the SP accessing his/her location takes place through processing and analyzing this collected mobility information. Privacy loss is then a function of the information disclosed from observing the user's mobility. In what follows, we define the mobility model, the privacy mechanisms of PR-LBS, and the cost function which enables the location–service exchange of PR-LBS as will be evident later.

5.1 Mobility Model

Topology: PR-LBS views the topology of the area as an unweighted and undirected graph $G = (Z, E)$, where

Table 1. The symbols table.

Symbol	Meaning
Z	set of zones in the area of interest
N	number of sessions
p	path traversed in a session
$l[p]$	length of traversed path in a session
$n[p]$	number of observations of path p
$P(p) = n[p]/N$	probability of traversing a path p
p_l	a single path p of length l
P_l	the set of paths of length l (all paths p_l)
$d_G(z_i, z_j)$	distance between two zones: z_i and z_j
$t_G(z_i, z_j)$	shortest time to travel between two zones

Z, the set of zones represents the nodes and E is the set of edges of the graph representing the transitions between neighboring zones. Each edge, e, is associated with the time, $t(e)$, the user takes to travel along it. In a typical public space, all zones are reachable from the entrance, making the graph connected. We can then define a path as the sequence of visited zones (of interest to the user) in the graph as: $p_l = \langle z_k \rangle_{z_k \in Z, z_k \neq z_{k+1}}$, where l is the path length (number of zones) and z_k is the k^{th} zone of the path. To count a zone as part of the path, the user must *visit* the zone and stay there for at least 30 seconds, not just passing through.

We define two functions in the graph G: the distance between two zones in the graph $d_G(z_i, z_j)$ is the length (number of edges) of the shortest path between z_i and z_j; and the time between two zones $t_G(z_i, z_j)$ as the shortest time it takes the user to travel between two zones (taken as the shortest path when the weights in the graph are considered as $t(e)$ instead of 1).

PR-LBS only releases the path defined above (p_l) or a variant thereof. Therefore, it hides the intra-zone as well as low-level mobility and only releases significant changes in the user's location (when visiting a new zone). As such, consecutive zones in the path need not be geographical neighbors; the path, as we define it, is not equivalent to the actually traversed path, but rather a part of it. For example, the user might have traveled along the zones A-B-C-D, but only spent time at A and C. PR-LBS releases the path (or a variant of) A-C.

Sessions: The user's mobility is broken down into *sessions*. In each session, the user enters the area, traverses a path, and then leaves; i.e., a session maps to one traversed path. We focus on the path as it embodies all the information about the user's mobility including the zones of interest, their priority and importance to the user, and other tracking information. The path starts at the beginning of a session and ends at the end of the session; subpaths do not count as independent paths. Therefore, each session will be associated with one path p of length $l[p]$.

We model the user's mobility as the probability distribution of a set of paths s/he traverses. PR-LBS populates this mobility model empirically based on the SP's observations (zones that PR-LBS revealed to the SP). After N sessions, the SP observes the user traversing a path p for $n[p]$ times. Each path in the mobility model is a distinct event; the probability of the user traversing each path (as observed by the SP), $P(p)$, is simply the count of the path divided by the number, N, of the user's sessions, $P(p) = \frac{n[p]}{N}$.

The set of paths of equal lengths ($P_l = \{p | l[p] = l\}$) forms a probability distribution: $\forall p \in P_l, P(p) = \frac{n[p]}{N}$. The probability of each path is the probability of the user following a path of the same length in a session. As some sessions will not have a path of length l, the probability distribution will include the event of the user not following a path of such length denoted by $P(\langle \phi_l \rangle) = 1 - \sum_{p \in P_l} P(p)$.

5.2 Private Location Release Mechanisms

PR-LBS protects the privacy of the user's mobility by anonymizing the traversed paths at runtime. PR-LBS has to guarantee an entire path's privacy while sequentially releasing zones along the said path, i.e., before it knows what the path is going to be. This is very different from most of the existing approaches that consider offline private publishing of mobility traces including the works by Rastogi and Nath [20], Abul *et al.* [21], Terrovitis and Mamoulis [22], and Chen *et al.* [23]. Moreover, adding noise, drawn from a distribution [24, 25], on the user's visited location does not apply in the indoor case. In most cases, the user's location is defined in terms of a zone, such as a UUID of an iBeacon, rather than a geographical location ($< x, y >$) so that noise drawn from some planar Laplacian distribution can not be added to a UUID value of an iBeacon.

The main privacy protection of PR-LBS comes from anonymizing the user's path, i.e., releasing a path, $path_{obs}$, instead of the actual traversed path. In particular, PR-LBS aims to provide (ϵ, d_m) differential privacy [18, 26–29] such that:

$$P(path_{obs}|path) \leq e^{\epsilon} P(path_{obs}|path'), \qquad (1)$$

where $d(path, path') \leq d_m$ and $P(path_{obs}|path_{tr})$ is the probability of observing $path_{obs}$ given the user traversed $path_{tr}$.

The criterion of Eq. (1) states that the privacy preserving mechanism releases a path, $path_{obs}$, (observed

by the SP), such that the probability of this path being the result of applying the privacy mechanism on the actually traversed path is indistinguishable (to an exponential factor) from that of applying the same mechanism on another path at most a distance d_m from the actual path. In other words, the SP, after observing $path_{obs}$, can not identify the user's actual path. The user's actual path is indistinguishable among the set of paths within a distance d_m from the user's actual path – we refer to this set as $path_{d_m}$.

The main challenge here is that PR-LBS can not treat the user's path as a series of zones devoid of any geographical significance. Blindly attempting to satisfy Eq.(1) will help the SP narrow down the search space by eliminating some implausible paths from $path_{d_m}$, given the released path. For example, a user enters a tunnel that can only be traversed in one direction: A-B-C-D. If PR-LBS releases any path of length 4, then the SP will directly infer the user's original path as A-B-C-D (the only plausible path in $path_{d_m}$ regardless of the value of d_m in this case).

This challenge arises from the fact that for a certain path $pa \in path_{d_m}$, $P(pa|path_{obs}) = 0$ (implausible given the observation) so that by Bayes' rule $P(path_{obs}|pa) = 0$, which violates the promised differential privacy guarantees. PR-LBS ensures that for $\forall pa \in path_{d_m}$, the probability $P(pa|path_{obs}) > 0$ so that the SP can not reduce the size of the search space after observing $path_{obs}$. PR-LBS's privacy preserving mechanism need not ensure the observed path to be plausible per-se, but any path in $path_{d_m}$ must plausibly be the actual path, given the released $path_{obs}$.

Recalling that a path is a sequence of zones associated with time, a plausible path is one which the time separating each two consecutive zones allows for a person to travel between them. While the SP has access to the area's map, PR-LBS relies on previously recorded user mobility to populate the graph G describing the layout. Each time PR-LBS observes a new transition, it adds the newly observed edge to the graph along with the travel time. Eventually, PR-LBS populates the graph and uses it to compute the travel time between any two zones in the graph.

5.2.1 Differential Privacy (D.P.) Mechanism

PR-LBS chooses to release the user's visited zone with a probability q_0 and chooses a zone at a distance of i with a probability $q_i = \alpha^i.q_0$ such that $\sum_i \alpha^i.q_o = 1$ and $i < d_m$, where d_m is the indistinguishability threshold. We define the distance between two equal-length paths ($d(path, path')$) as the edit distance with a non-negative weight. The only operation we consider in the edit distance is substitution so that the weight/cost of each substitution is the distance ($d_G(z_1, z_2)$) between the two substituted zones. For example, the distance between two paths $A-B-C-D$ and $A-E-C-F$ is: 0 (cost of sub A with A) + $d_G(B, E)$ (cost of sub B with E) + 0 (cost of sub C with C)+ $d_G(D, F)$(cost of sub D with F). Since the weight between two zones is symmetric ($d_G(z_1, z_2) = d_G(z_2, z_1)$), the distance between two paths satisfies the axioms of a metric.

So, this mechanism achieves differential privacy such that (proof in Appendix A):

$$q_0 \leq \frac{1}{\sum \alpha^i} \; s.t. \; \alpha \geq \frac{|Z|_{d_m}}{e^{\epsilon/m}}, i < d_m, \tag{2}$$

where $|Z|_{d_m}$ is the number of zones within a distance d_m of the user's visited zones.

When the user moves from a zone za_l to za_{l+1} (with a travel time ta), s/he would have traversed a path pa of length $l+1$ so far. The D.P. mechanism releases a zone z' instead of za_{l+1} according to probability distribution described above. At the same time, PR-LBS keeps track of $path_{d_m}$, the set of paths of equal length of pa and of a distance less than d_m from pa. For each path p_r (comprised of zones zr_i) in $path_{d_m}$, PR-LBS estimates the travel time of the transition from zr_l to zr_{l+1} as $t_r = t_G(zr_l, zr_{l+1})$. If there is at least one path of $path_{d_m}$ where $t_r >> ta$, then PR-LBS hides za_{l+1} completely (and does not release any anonymized zone). Therefore, PR-LBS avoids releasing a path to the SP that violates the indistinguishability criterion. At the end of Section 7, we will show that PR-LBS can effectively distort the distribution of the zone visit time for privacy-oriented users. This prevents the SP from utilizing timing information to infer more probable paths from the set $path_{d_m}$.

The value of d_m controls the trade-off between privacy and utility. A lower value of d_m will allow the privacy criterion to be more relaxed (a higher value of q) so that the observed path will be closer to the actual path, thus becoming indistinguishable among a smaller set of paths.

Learning: While PR-LBS is populating the graph, it can not apply the above mechanism as it will not have a full view of the area's topology (a list of zones without transitions). In such a case, it applies a variant of the D.P. mechanism. In this variant, PR-LBS releases the user's visited zone, z_v, with a probability q or any other zone $z \in Z \setminus z_v$ with a probability $1-q$. We define

the distance, $d(path, path')$, between two paths $path$ and $path'$ as the edit distance with weight 1 (=the number of different zones between two paths). This mechanism achieves (ϵ, d_m) differential privacy (proof in Appendix A), where

$$q \leq \frac{e^{\epsilon/d_m}}{|Z| - 1 + e^{\epsilon/d_m}}. \quad (3)$$

This mechanism is not very efficient in terms of the privacy–utility tradeoff; it treats all the other zones as being equidistant to the current zone. Once the full topology of the graph is known, PR-LBS applies the full D.P. mechanism which exhibits a finer-grained privacy–utility tradeoff by providing indistinguishability over topologically close paths.

5.2.2 Anonymity Set (A.S.) Mechanism

In some scenarios, PR-LBS runs on a device with no capability of feeding the SP app with a fake zone instead of the user's visited zone; it can control whether to release or hide the currently visited zone. Therefore, PR-LBS can not apply the D.P. mechanism described above.

Instead, PR-LBS resorts to the A.S. mechanism that releases the user's visited zone with a probability q and hides it with a probability $1-q$. This mechanism can not provide differential privacy guarantees since there will always be a path such that the observation probability will be 0. For example, if PR-LBS releases the path $p_{obs} =< a, b >$, then the probability of observing this path given the user traversed $< c, a, d >$ is 0. The expression of Eq. (1) can not be satisfied. Therefore, we focus on another privacy indicator which is the size of the anonymity set. The anonymity set is defined as the set of paths from which the released path could have possibly resulted, i.e., $\forall p | P(path_{obs}|p) > 0$. In appendix A, we derive the expected size of the anonymity set for a traversed path of length m as:

$$E(S) = \sum_{k=0}^{m} \sum_{r=0}^{k} \binom{k}{r}^2 q^r (1-q)^{k-r} \frac{(|Z| - r)!}{(|Z| - k)!} \quad (4)$$

As evident from the expression of Eq. (4), the value of q controls the uncertainty at the SP side. For instance, when $q = 0$, the value of $E(S)$ assumes the maximum value since the adversary will not observe any of the user's mobility. The mobility inside a zone and between two released zones (z_i and z_j in this case) is always hidden. During this gap, the user could have spent time at z_i or in one or more zones in between (on path p_l). The SP can not definitely decide whether the user actually visited a zone in between (on path p_l) or spent the entire time between at z_i.

5.2.3 Path Diversity

Both D.P. and A.S. mechanisms rely on hiding the user's visited path within an anonymity set of other paths. The size of this set provides the privacy guarantees to the user and is mainly controlled by the number of zones in the area and the possible transitions between these zones as recorded by previous user mobility. It is very important to note here that while mobility can be restricted in an indoor case, our definition of a zone visit relaxes this restriction. In particular, in an indoor space, the graph depicting the area's topology is connected so that every zone is reachable from any other zone.

Since consecutive zones in a path are not geographically adjacent, the number of zones reachable from the currently visited zone is not restricted to neighboring zones, but by their feasible transitions. PR-LBS copes with the issue of limited feasible transitions, which takes place at the bootstrapping stage, by maintaining the size of the anonymity set for both mechanisms. When the size of the anonymity set is small, PR-LBS hides the currently visited zone.

5.3 Information Disclosure

In what follows, we analyze the (privacy) cost incurred from PR-LBS's release of a path to the SP (even if it is anonymized). Any observation (a zone visit) will necessarily change the probability distribution spanning the mobility model. The amount of change introduced to the mobility model is what we attempt to quantify. Even when PR-LBS releases an anonymized path, this path will still carry information of the actual mobility.

We first quantify the information disclosure (alternatively leak) for the entire user's path after observing a new zone visit and then state our information leak model for the specific zone visit.

5.3.1 Per-path Criterion

To quantify the information leak for observing a path, we follow the lead of Miklau and Suciu [30] by considering a metric of positive information disclosure. We focus on the improvement of the probability of the

user traversing a certain path as being indicative of the amount of information released:

$$lk(s,v) = \sup_{s} \frac{P(S=s|v) - P(S=s)}{P(S=s)}. \quad (5)$$

In Eq. (5), $P(S=s)$ is the *prior* probability distribution of a secret s that the adversary attempts to identify, v is the observation, and $P(S=s|v)$ is the probability distribution of S after observing v.

In our setting, the secret that the adversary wants to unravel is the user's probability distribution of the paths traversed. When PR-LBS is about to release a new zone z_l to the SP, after releasing a path p_{l-1}, it estimates the information leak of the total observed path being $p_l = \langle p_{l-1}, z_l \rangle$. The information disclosure considers the improvement of the SP's observation probability of the user traversing a path as being indicative of the amount of information released. If the observation of a path does not improve the adversary's knowledge, it leaks little/no information about the user to the SP (the SP already expects the user to traverse such a path), and vice versa.

If the user has visited the area N times (number of sessions), out of which s/he traversed the path p_l for $n[p_l]$ times, then the per-path information leak is (see the derivation in Appendix A):

$$lk(p_l, z) = \frac{1-a}{a(N+1)},\ a = \frac{n[p_l]}{N}. \quad (6)$$

5.3.2 Per-zone Criterion

We can rewrite the leak function of Eq. (6) to to represent the information leak in bits as: $leak(p_l, z) = \log_2\left(lk(p_l, z) + 1\right) = \log_2\left(\frac{N(n[p_l]+1)}{n[p_l](N+1)}\right)$.

It is worth noting that $leak(p_l, z)$ represents the improvement of the observer's knowledge of traversing a path, p_l, directly after the observation of z. Having visited N sessions, of which a path p_l has been traversed $n[p_l]$ times, the probability of visiting path p_l is originally $P(p_l) = n[p_l]/N$. For the $(N+1)^{th}$ session, if p_l is traversed, it will be the only path (of the user's mobility model) experiencing a positive information disclosure as $P(p_l|z)$ will be $(n[p_l]+1)/(N+1)$, which represents $P(S=s|v)$ of Eq. (5). It is straightforward to show that for $N > n[p_l]$ (which is always the case since N is the total number of sessions), $\frac{n[p_l]+1}{N+1} > \frac{n[p_l]}{N}$ so that the information disclosure will always be positive. Therefore, applying the logarithm to compute $leak(p_l, z)$ is always feasible.

For $N > 0$, $n[p_l] > 0$ and $n[p_l] \leq N$, the value of $leak(pv_n, z)$ varies between 0 (minimum leak) and 1 (maximum leak). In the initial case where no mobility has been observed about the user ($N = 0$ or $n[p_l] = 0$), we associate $leak(p_l, z)$ with the maximum value of 1. We define the information leak per observed zone as the difference between the leaks resulting from the paths p_l and p_{l-1}:

$$\begin{aligned} leak(z) &= leak(p_l, z) - leak(p_{l-1}, z) \\ &= \log_2\left(\frac{n[p_{l-1}](n[p_l]+1)}{n[p_l](n[p_{l-1}]+1)}\right). \end{aligned} \quad (7)$$

A closer look at Eq. (7) reveals that the leak per zone is the leak defined by the conditional probability distribution $P(z_l|p_{l-1})$ which is $\frac{n[p_l]}{n[p_{l-1}]}$. Since PR-LBS considers the user mobility one zone at a time, by the time the user reached z_l, the entire path p_{l-1} must have been observed by the SP. Hence, $leak(z_l)$ focuses on the additional leak incurred from releasing z_l given that the SP has already observed the user's path p_{l-1}. Appendix B provides an example of how PR-LBS computes the privacy cost of a traversed path.

This information leak is especially crucial for the case of the A.S. mechanism which will release raw location data. We rely on the information disclosure as a cost metric to cap the additional knowledge leaked about the user to the service provider.

Finally, the information leak as defined in Eq. 6 offers a nice property. If the value of N is large enough then the information leak from observing path p_l can be approximated by $\frac{1}{n[p_l]}$. This implies that the first observations of a path will have higher leaks compared to future observations. On the other hand, when the value of N is low, there is always going to be a leak of information, as the probabilities of following paths will be changing more abruptly. For a fixed N, a lower probability $P(p_l)$ of traversing a path will always lead to a higher information leak. Individuals tend to perceive behaviors with low probability as being more private because they indicate unexpected (thus conspicuous) behaviors [31, 32].

6 PR-LBS

In what follows, we describe PR-LBS, its different components, and their interactions.

6.1 The Location-Service Exchange

We model the interactions between the user and the SP as a repeated play model [6] composed of the user

Table 2. The experts utilized by PR-LBS.

Expert	Advice
First	Hides location all time
Second	Release location according to privacy mechanism
Third	Release location all time

(player) and the SP (opponent). The user, with PR-LBS acting on his/her behalf, chooses one of three actions: *hiding*, *releasing*, or *anonymizing* the location. In return, the SP pushes a service with varying value.

We rely on the SEA algorithm [7] to decide on which action to take at each phase (when visiting a new zone). In this model, the player has access to a set of experts each offering an advice for the action to take at each interaction. The algorithm is akin to reinforcement learning and is based on a combination of exploration and exploitation. The exploration phase will enable the player to learn the opponent's response to the different actions, while the exploitation phase enables the player to follow the expert's advice who has accumulated the highest average utility.

SEA has some nice properties that make it suitable for our context. First, it assumes a non-oblivious opponent whose actions might (or might not) depend on the player's actions (as in our case) as opposed to minimum regret algorithms [6]. SEA, also, avoids being short-sighted in its objective and instead focuses on the asymptotic behavior of the player. Finally, it can achieve a stricter bound if the opponent is assumed flexible. A flexible opponent is one who forgets the player's actions after a while. Analytics and advertisement servers constitute a relevant example, their user recommendations are usually based on recently observed behavior.

The user–SP interaction takes place when PR-LBS detects the user has visited a new zone. PR-LBS has a set of three experts with each recommending an action to follow as defined in Table 2.

The mechanism invoked by the second expert of the exchange module will depend on the capabilities available to PR-LBS. If PR-LBS can change the zone reported to the SP's app, then it can rely on the D.P. mechanism (while running in device-based mode on a rooted device or in infrastructure-based mode). Otherwise, the second expert uses the A.S. mechanism when PR-LBS can only enable/disable location release (while running in device-based mode on an unrooted device).

The player can designate an interaction as either an exploration or exploitation stage according to a biased probability distribution. Specifically, PR-LBS designates the i^th interaction as an exploration stage with a probability $1/i$ and as an exploitation stage with probability equal to $1 - 1/i$. In the exploration stage, the player chooses one expert at random. This ensures that every expert is sampled infinitely many times. In the exploitation stage, the player simply chooses to follow the advice of the expert with the highest accumulated average utility. At the earlier interactions, where the value of i is small, PR-LBS chooses exploration with a higher probability as to learn the utilities of the different experts. With more interactions, the behavior of PR-LBS stabilizes and it chooses exploitation with a higher probability as it would have accumulated enough average utility for each expert.

After PR-LBS chooses an expert, it follows its advice for the coming interaction between the user and SP. Every interaction involves deciding on an action (based on the expert's advice), computing the privacy cost of the performed action, estimating the service value of the whole interaction, and computing the utility resulting from the interaction (a function of the cost and reward). After the interaction ends, PR-LBS updates the average utility for the chosen expert. PR-LBS, then, chooses another expert for the next interaction.

6.2 Utility Estimator

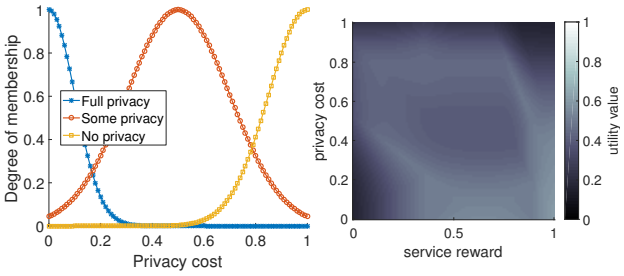
Deriving a utility estimate from the privacy and QoS estimates is not straightforward; it is like comparing apples to oranges. Besides, the utility function is subjective as it depends on the user's perception. A privacy-oriented user will suffer lower utility if more location samples are released, while a service-oriented user will suffer lower utility if s/he does not receive services.

In PR-LBS, the user defines a privacy profile which indicates his/her tradeoff between the cost of releasing location and the benefit from the received service. The utility estimator module converts this "high-level" profile to a utility function that maps the privacy cost and service quality pair to a value between 0 (no utility) and 1 (full utility). Table 3 shows an example of the privacy profile of a privacy-oriented user (from our survey). In the survey, we asked respondents to fill in their privacy profile through a table similar to Table 3. Each entry specifies the respondent's satisfaction (0 – not satisfied, 1 – somehow satisfied, and 2 – fully satisfied) value for each of the privacy and service combination. It is clear how this respondent favors curbing location sharing.

Given a privacy profile (one that looks like Table 3), the utility estimator module converts it to a numerical function. The resulting function takes two inputs: pri-

Table 3. The privacy profile of a privacy-oriented user.

	No Service	Some Service	Full Service
No Privacy	**0**	**0**	**1**
Some Privacy	**0**	**0**	**1**
Full Privacy	**1**	**1**	**2**



(a) Privacy membership function

(b) The resulting utility function.

Fig. 4. Utility estimator illustration.

vacy cost ($leak(z_i)$) and estimated service quality values ($serv_i$); it returns an output which is the utility value such that $utility = f(leak(z_i), serv_i)$. As the concepts of the privacy profiles (privacy, QoS, utility) are defined in qualitative (or humanistic) terms, a fuzzy inference system (FIS) [33] is thus suitable to derive the utility function out of these values.

We rely on a Mamdani-type fuzzy inference system; such a system has two main components: rules (defining relationship between inputs and outputs) and membership functions. In our context, the rules are defined as per the privacy profile (similar to Table 3). For example, the top left entry of the table defines this rule: *No Privacy* ***AND*** *No Service* $\implies$ *Not satisfied*.

The membership function defines the value's relevance to the property it is claiming over a domain. For example, instead of defining a hard threshold between what is a low threat path and a high threat one, one can define softer thresholds, as evident from Fig. 4a. A path definitely poses a low threat (full privacy) when $leak(p) = 0$, as the value of 0 has a membership of 1 in the low function. The membership value decreases gradually as the privacy value increases until it hits 20%. Similar membership functions can be defined for medium, high threat. A similar logic applies for the service and utility values.

The second step in the fuzzy inference system is mapping the inputs to the output; this is achieved through the fuzzy operators, which results in a fuzzy utility value. The final stage is the conversion of a fuzzy utility value to a crisp output to determine the actual utility to feed the SEA algorithm through the defuzzification step. Ultimately, the FIS will result in a continuous and smooth 2D function mapping both the privacy and service values to a utility value. Fig. 4b shows the final utility function for the privacy profile of Table 3. The utility takes its maximum value at maximum privacy and reward levels. The utility, smoothly, decreases as the privacy level or reward decrease.

6.3 QoS Analyzer

Measuring the quality of service the user received from interactions with the SP's app is essential to the operation of PR-LBS. Market research literature has a wealth of studies that analyze the user's retail app usage and its effect on user satisfaction and purchases [34–42]. This literature leads to the following conclusions regarding retail apps:

1. Retail apps rely heavily on push notifications to communicate retail services to users, which we confirmed from our analysis of multiple retail apps. In 2014, more than 80% of the notifications pushed by the retail apps were consumed by users [39, 43].
2. Continued app usage and interaction inside the store (during shopping) directly relates with user's satisfaction during the shopping experience [34–38].
3. Higher retail app usage rate (on-screen time) during shopping is correlated with more brick-and-mortar store visits, longer shopping visits, and increased purchase rates [40–42].

Consistent with market research literature, we utilize user interaction with the SP's app as an indicator of the user's satisfaction with the services provided by the app. An interaction with a typical retail app takes place in three stages: (1) the app pushes a service to the user through a notification, (2) the user consumes the notification (by checking and reading it), and (3) the user opens and interacts with the SP's app. To measure user interaction with the SP's app, PR-LBS observes if the user consumed a push notification from the app and measures the time s/he interacted with the app.

PR-LBS monitors the level of user–app interactions through the collector module to compute the QoS metric: $serv_i$. It observes the time the user spent interacting with the SP's app, denoted by $time_{foreground}$, that was preceded by a consumed push notification from the SP's app. Particularly, if the user consumed a push notification by opening the SP's app, then PR-LBS measures

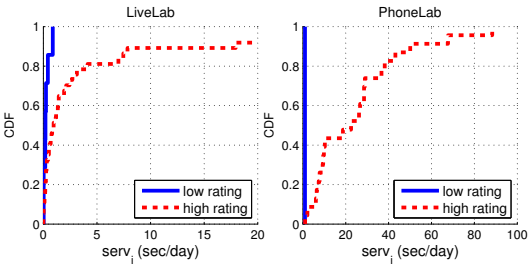


Fig. 5. The distribution of QoS metric for the LiveLab (left) and PhoneLab (right) datasets.

the fraction of time the user spent actively interacting with the app as:

$$serv_i = \frac{time_{foreground}}{time_{z_i - z_j}}, \tag{8}$$

where $time_{z_i - z_j}$ denotes the time spent in the last zone(z_i), before moving to the new zone(z_j).

To further assess whether $serv_i$ is a good indicator of user satisfaction with retail apps, we analyzed retail app usage data from two datasets: the LiveLab dataset of Rice University [44] and our dataset of participants whom we recruited from PhoneLab [45]. The Livelab dataset contains the app usage data, along with other data, for 34 iPhone users over a 12–18 month period. Our dataset has app usage data for 95 Android users over 4 months. We identified retail apps from their app category in either Google Play or iTunes. For every retail app and user combination, we calculated $serv_i$ as the app total usage time (in seconds) normalized by the installation time (in days). We then associated every app session with its average rating on the app store. We categorized apps into two categories, in terms of rating: low (rating ≤ 2.5) and high (rating ≥ 4).

Fig. 5 plots the distribution of the $serv_i$ for each of the two categories (low and high) for both datasets. There is a large discrepancy between the high and low distributions suggesting that highly rated apps tend to enjoy higher usage rates. For the apps with a low rating, 80% of the apps are used at most 0.1 sec/day, while 80% of the highly used apps have more usage rate than that. The discrepancies of values between both datasets relates to the duration of each dataset (Livelab is much longer than PhoneLab). It is clear that $serv_i$ correlates with the average user rating of the app at the app store (iTunes/Google Play) – app user rating acts as an indicator of user satisfaction of its service [46].

The proposed reward metric offers several advantages in terms of practicality and usability. **First**, it limits interactions with the user. Existing methods that rely on surveys to measure user satisfaction do not apply in our context. PR-LBS needs to measure user feedback at a "micro-scale" as the user is moving from one zone to the other, not after the visit is completed. It is impractical to continuously ask the user for feedback about the received service as s/he moves from one zone to the other. **Second**, the reward metric is app-agnostic; it does not require specific knowledge of the semantics of the service provider's app. Otherwise, PR-LBS has to tailor its estimation methodology of the service rewards to every service provider app, which is impractical. **Finally**, $serv_i$ is practical to measure as it can be extracted from user-level information on the user's device. It requires neither intercepting network traffic nor instrumenting the user's mobile operating system.

7 Implementation and Evaluation

We now present the implementation of PR-LBS in device mode and the evaluation of its effectiveness.

7.1 Implementation

We implement the device mode of PR-LBS as a standalone Android (4.4) app, which is compatible with beacon-based localization. Fig. 6 shows the architecture of the PR-LBS app. PR-LBS runs as a background service that detects if the user is visiting a place where localization is deployed. When the user starts the visit, PR-LBS prompts him/her to set the privacy preferences and executes the logic of Fig. 6. The privacy preferences include setting the privacy level and the privacy profile of Fig 8. The privacy level ($priv_{lvl}$) is a slider between no privacy $priv_{lvl} = 0$ and full privacy ($priv_{lvl} = 1$) that sets the parameters of the privacy mechanisms as such:

D.P. Mechanism: if the maximum distance (length of shortest path) between any to zones in the area is d, then $d_m = priv_{lvl}.m.d$. When its learning variant is running, then d_m is set as: $d_m = priv_{lvl}.m$, where m is the average path length.

A.S Mechanism: the probability controlling the release of location is simply set as $q = 1 - priv_{lvl}$.

BLE Scanner: PR-LBS utilizes Android's Bluetooth Low Energy (BLE) APIs to scan regularly for BLE beacons. During the scan duration, PR-LBS receives advertisements from multiple beacons. It decides on the beacon with the lowest power attenuation as the closest to the user. It extracts the identifying fields from the beacon advertisement to maps a zone *id*. If the new

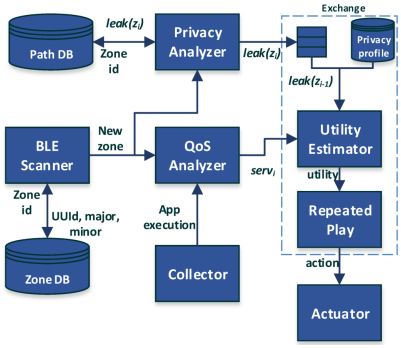


Fig. 6. The architecture of PR-LBS in device mode.

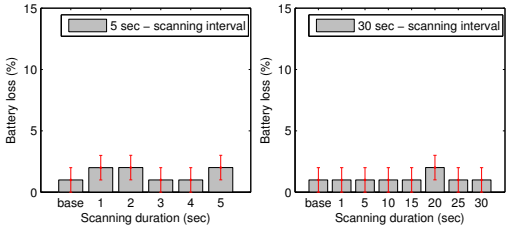


Fig. 7. The energy consumption by PR-LBS.

Fig. 8. The UI to input the privacy profile in PR-LBS.

zone is different from the last detected zone, the scanner alerts the QoS and Privacy analyzers. PR-LBS uses the scanned zones to populate the topology graph.

Collector: This module records app execution events and keeps track of the time the user spent for interacting with the service provider's app. As Android does not provide a public API for this purpose, the collector module frequently polls (once a second) the running tasks to find which apps the user is currently running in the foreground. It also runs an an `Accessibility Service` to intercept the SP app's notifications and the resulting user actions. Whenever a new zone is detected, the collector passes this information to the QoS analyzer that calculates the QoS metric.

Actuator: The actuator is responsible for performing the action decided by the exchange module. The action could be *hide*, *release*, or *anonymize* the visited zone. While running on an unrooted device and with the absence of any other support, the only actions available are to release or hide the currently visited zone (A.S. mechanism). In such case, PR-LBS uses the Android's Bluetooth Admin permission to globally control Bluetooth scanning on the device. This will prevent the service provider's app (and potentially other apps) from tracking the users. On the other hand, when running on a rooted device, PR-LBS will be able to apply the *anonymize* action. We instrument the Bluetooth scanning function in Android's framework to so that PR-LBS changes the Bluetooth Low Energy scan results that the SP's app will receive. We are currently exploring using the user's smartwatch to advertise dummy beacons to anonymize the currently visited zone which will not require rooting the user's smartphone.

7.2 App-Based Evaluation

We evaluate the energy overhead of PR-LBS when it runs on Samsung Galaxy S5 (Android 4.4.4). We deployed a set of iBeacons in a lab environment with one iBeacon (closest to the device) continuously changing its identifiers, making PR-LBS believe a new zone was detected upon each BLE scan. Each new zone event triggers PR-LBS to execute its components (Fig. 6). Hence, the frequency of detecting new zones along with the scanning interval (frequency) and duration (length of the scan) determine PR-LBS's energy consumption. We report on the battery energy consumption (which includes the full operation of PR-LBS with all of its components) in Fig. 7 when PR-LBS runs under two scanning intervals: 5s and 30s. For each scanning interval, we vary the scan duration to study its effect as shown in Fig. 7. We also compare the battery loss with the base case, with PR-LBS turned off. We run all the experiments for 10 minutes with the screen fully lit while turning off background apps, Wi-Fi, and data connections. It is clear from Fig. 7 that PR-LBS incurs limited energy overhead since PR-LBS is a lightweight app that incurs little processing overhead.

We also test the usability of the privacy profile input interface (Fig. 8). We deployed PR-LBS on Google Play and asked 100 participants (recruited over Amazon Mechanical Turk) to test the app and answer a set of questions. We paid each participant $1 and the average time for tasks completion was 7 minutes. When the participant completed interaction with the app, it displayed a special code to input in the survey to ensure completion of the required task. 93% of the participants considered that the interface is easy to use and 85% of them indicated that it is clear.

Table 4. The eight datasets used for the evaluation.

Dataset	Location	Duration (days)	# zones	# users
HOPE 2008 [47]	Hotel	3	21	1273
HOPE 2010 [48]	Hotel	3	26	1119
State Fair [49]	Fair	5	32	19
Orlando [49]	Theme park	61	44	41
NCSU [49]	Campus	80	49	35
KAIST [49]	Campus	78	44	92
Walmart	Retailer	-	29	93
Nordstrom	Retailer	-	34	76

7.3 Trace-Based Evaluation

Datasets: We utilize 6 datasets from the CRAWDAD data repository to evaluate PR-LBS. These datasets represent the mobility of individuals in public constrained spaces that PR-LBS operates in. We also utilize two other datasets that belong to respondents from our survey. In both surveys (Walmart and Nordstrom), we displayed a map of the store with labeled zones. We asked each participant to trace the path s/he traversed during the last visit. Table 4 shows a list of the datasets.

To evaluate PR-LBS, we transform each dataset into a stream of location samples. PR-LBS processes every location sample regardless of whether it came from the real world or a location trace, which indicates that our evaluation is representative of PR-LBS's operation in the real world. We further partitioned every stream into sessions or paths. The last two datasets, Walmart and Nordstrom, had one path per user and no time information associated with the location trace.

Scenarios: We simulated four classes of SPs that reward the user for sharing location information differently, while not rewarding for hiding location. In our model, the SP will offer the user a service with a reward value ($serv_i$) between 0 and 1. This abstracts both the SP and the QoS analyzer module. The SP classes are:

1. *None*: No reward for the user.
2. *Low*: Low reward (below 0.3) for sharing.
3. *Med*: Medium reward ($\geq$ 0.3 and $\leq$0.8) for sharing.
4. *High*: High reward (higher than 0.8) for sharing.

We consider the three privacy profiles defined in Section 3: service-oriented, neutral, and privacy-oriented, which we constructed based on our survey results. Each respondent filled a table exactly like Table 3 where each response corresponds to the privacy profile of the respondent. To generate the three privacy profiles, we average the profiles of each (as defined in Section 3) respondent (in the three profiles) and then round the values to the nearest integer.

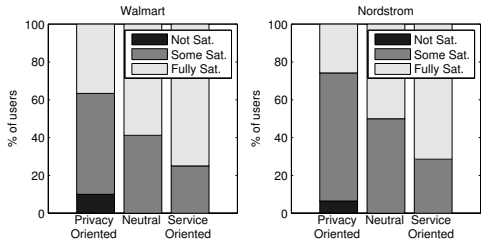


Fig. 9. User satisfaction distribution with high service.

We execute PR-LBS for each user in each dataset for each scenario (service class and privacy profile combination). At the end of each run, PR-LBS would have released some of the user's mobility. Our evaluation is based on comparing, for each user and in each scenario, the released paths with their original counterparts.

PR-LBS Feasibility: In Section 5.2.3, we indicate that even in a constrained indoor area, the number of feasible transitions (not necessarily from geographically neighboring nodes) is what matters for PR-LBS's operation. For the eight mobility datasets of Table 4, the number of geographical transitions is indeed limited by the area's topography. Nevertheless, the size of the set of feasible transitions (measured from user mobility) is near perfect for all the datasets. In particular, the "NCSU" dataset has 1912 feasible transitions out of 2352 possible ones, "State Fair" has 992 out of 992, "Orlando" has 1724 out of 1892, "KAIST" has 1892 out of 1892, "Walmart" has 812 out of 812, "Nordstrom" has 1122 out of 1122, "Hope 2008" has 420 out of 420 and "HOPE 2010" has 650 out of 650. The number of possible transitions in an area is simply $|Z|.(|Z|-1)$, where $|Z|$ is the total number of zones.

Since PR-LBS considers only the visited zones as composing a path, it is able to overcome constraints in an area's topology. It exploits the larger set of feasible transitions to provide a larger anonymity set for both the D.P. and A.S. mechanisms. PR-LBS need not hide zones because of infeasible transitions, which maintains utility for the user and service providers.

User satisfaction: To study whether PR-LBS matches user expectations, we execute PR-LBS over each path of both Walmart and Nordstrom datasets with the different SP models (low, medium, and high). At the end of each run, we capture the path that PR-LBS releases and we compute the privacy metric ($leak(path)$) according to Section 5. We also compute the average service value received per zone. We end up, for each user, with the service value that the user received and privacy loss experienced.

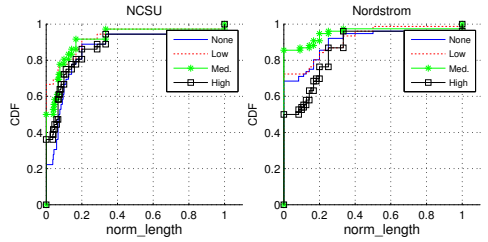


Fig. 10. *norm_length* for a privacy-oriented profile.

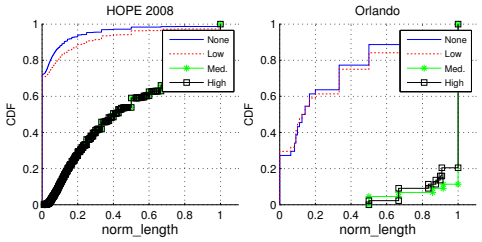


Fig. 11. *norm_length* for a a service-oriented profile.

Table 5. The utility metrics description.

Metric	Description
first	**Portion of paths with correctly identified start zones.**
last	**Portion of paths with correctly identified end zones.**
dwellarea	**Portion of zones with accurate estimate of percentage time spent over all the users.**
transitions	**Portion of correctly identified zone transitions.**
retention	**Portion of zones with accurate estimate of retention (number of customers in a zone at a one time).**
dwelltime	**The portion of the zones in the area with accurate estimate of the dwell time. The dwell time of a zone is the average time spent at the zone.**

We follow the same procedure as Section 6.2 to build a utility function for each user in both the Walmart and Nordstrom datasets from the table containing their privacy profile. Recall that these profiles are nothing else than a mapping between a privacy – service pair to a utility metric (reflecting satisfaction). We then use the privacy and service values of each path as inputs to each user's profile to estimate his/her satisfaction. We round the output to the nearest integer reflecting three satisfaction levels: "not satisfied at all", "somehow satisfied", and "fully satisfied".

Fig. 9 shows the user satisfaction distribution for each privacy profile and for the high service level. The percentage of unsatisfied users is close to 0 in all the situations. Also, service-oriented users (and neutral users to a lesser degree) tend to be more satisfied than other users because they accommodate more location sharing. Although not shown due to space limitation, service levels correlate with user satisfaction for all three profiles; the higher the service is the more satisfied the users are.

Privacy Protection: We study PR-LBS's effect on protecting users' privacy through *norm_length*: the number of zones that PR-LBS released and the user actually visited divided by the total length of the original path. This metric indicates how much of the user's actual mobility information has been released.

Fig. 10 shows the distribution of *norm_length* for users with a privacy-oriented profile for the "NCSU" and "Nordstrom" datasets (other datasets show similar results, but are omitted due to space limitation). It is evident that PR-LBS curbs location sharing for privacy-oriented users with 60% of the paths hidden completely regardless of the service level and even when $p = 1$. PR-LBS shares some non-private location data (according to the cost metric) as part of its exploration stage.

Fig. 11 shows the distribution of *norm_length* for service-oriented users in the "HOPE 2008" and "Orlando" datasets. The amount of sharing is higher than that for privacy-aware users and roughly corresponds to the SP's service level. There is one caveat: sharing is mobility-dependent. PR-LBS hides more of the user's location for the HOPE 2010 dataset than the State Fair dataset. In the State Fair dataset, the environment is more constrained. Individuals exhibited less diverse paths and the portion of paths leaking information according to the metric of Section 5 were negligible. The mobility in "HOPE 2010" dataset is diverse with most of the paths exhibiting a high privacy threat. In such a case, even if the rewards provided by the SP are high, PR-LBS reduces sharing to protect the user's privacy.

Service Provider Perspective: Currently, SPs focus solely on aggregate analysis in an effort to comfort users. PR-LBS improves on the status-quo by enabling SPs with the capability to perform personalized analysis. In what follows, we evaluate the SP's utility using seven metrics [50, 51] as defined in Table 5. The closer these metrics are to 1.0, the higher is the utility that the SP enjoys. To model the SP's service level, we relied on the service value estimates of Fig. 5 from our Phonelab dataset, instead of using the synthetic values. We normalized the service values to fall between 0 and 1. Fig. 12 shows the metrics for each dataset and user privacy profile for these realistic service values.

First, the performance of PR-LBS is consistent among the different datasets which shows it can adapt to different environments and settings. **Second**, PR-LBS ensures a decent utility level even a significant portion of the users' mobility is not shared. Fig. 11 (left) shows that PR-LBS hides a significant portion of users' paths to protect their privacy for HOPE 2010 dataset (the re-

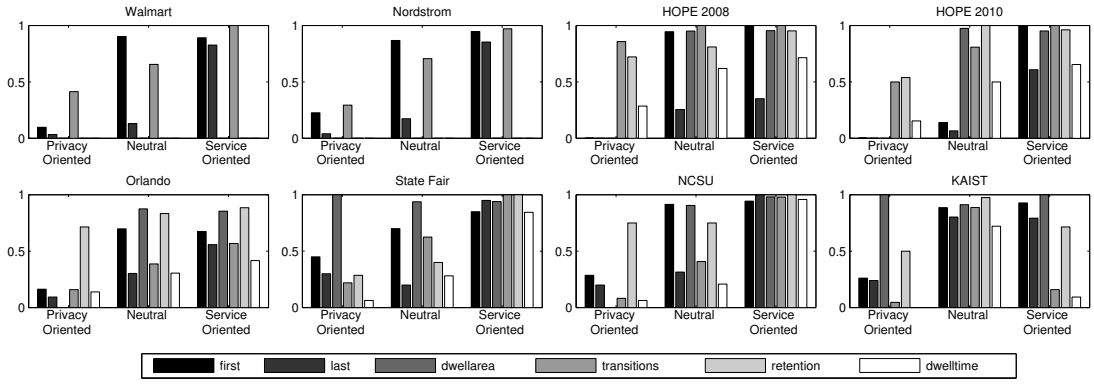


Fig. 12. Utility metrics for realistic service levels. Some metrics in Walmart and Nordstrom datasets are not available.

sults are similar to HOPE 2008). Still, the utility metrics are fairly accurate as they represent an aggregate of users' mobility. **Third**, as shown in Figs 14 and 13 (in Appendix C for space considerations), PR-LBS adapts sharing to the service level; when the server is more rewarding, PR-LBS will react by sharing more of the user's mobility, and vice versa. PR-LBS can incentivize the SP to reward the user with better service as it reflects with improved utility. The SP's utility decreases when it provides lower service to the users.

Finally, PR-LBS effectively protects the privacy of the privacy-oriented users by releasing fewer data about them. More importantly, the SP's accuracy in deciding the user's dwell time is always less than 10%. This indicates that the SP can not use the dwell time distribution for these users to identify possibly hidden zones from the timing information in the observed path.

8 Limitations

Lack of User Feedback: To achieve a usable and practical user experience, we made a conscious decision not to require user feedback regarding privacy decisions and rewards estimation. While PR-LBS attempts to estimate the privacy cost objectively and provide privacy guarantees, it does not capture the user's stance towards hiding or revealing every visited zone. Similarly, the reward metric of Section 6.3 might not be very accurate in describing the user's satisfaction with the provided service. In the future, we will investigate mechanisms to incorporate user feedback, in a usable manner, to improve the privacy and service estimations.

Evaluation Methodology: Our evaluation methodology suffers an inherent limitation. It assumes that user's privacy preferences are stationary, while they are prone to change if the user is presented with information about the SP's access to his/her location. Although our survey (see Section 3) results indicate the respondents' privacy preferences did not change before and after we informed them about retailers accessing their location, we believe this part warrants additional investigation in the future.

9 Conclusion

In this paper, we have designed, implemented, and evaluated PR-LBS, a system that balances between privacy and rewards in an indoor environment. It creates a win–win situation for both the users and service providers. PR-LBS packs in two mechanisms for private location release in indoor environments as well as a novel privacy criterion to estimate the cost of sharing location. It subjects the user's mobility to a location–service exchange that is based on the repeated play model to ensure the users are rewarded for sharing some aspects of their mobility. Our evaluations show that PR-LBS is easy to deploy, has low energy overhead, is usable, effectively remedies the user's concerns, and does not affect the SP's utility. In future, we would like to conduct a field study of PR-LBS's device-based prototype. We also want to explore options to providing APIs for the SPs to access aggregate information privately without releasing any raw location data.

Acknowledgments

We would like to thank the anonymous reviewers for constructive suggestions. The work reported in this paper was supported in part by the NSF under grants CNS-1114837 and CNS-1505785.

References

[1] S. Sen, J. Lee, K.-H. Kim, and P. Congdon, "Avoiding multipath to revive inbuilding wifi localization," in *Proceeding of MobiSys '13*, 2013, pp. 249–262. [Online]. Available: http://doi.acm.org/10.1145/2462456.2464463

[2] A. Martin, "Nordstrom no longer tracking customer phones," http://cbsloc.al/1JIYNIR, May 2013.

[3] Future of Privacy Forum, "Mobile Location Analytics Code of Conduct," http://www.futureofprivacy.org/wp-content/uploads/10.22.13-FINAL-MLA-Code.pdf.

[4] L. Privat, "U.S. consumers reject in-store tracking said survey," http://www.opinionlab.com/media_coverage/u-s-consumers-reject-in-store-tracking-said-survey/.

[5] H. Xu, H.-H. Teo, B. Tan, and R. Agarwal, "The role of push-pull technology in privacy calculus: The case of location-based services," *J. Manage. Inf. Syst.*, vol. 26, no. 3, pp. 135–174, Dec. 2009. [Online]. Available: http://dx.doi.org/10.2753/MIS0742-1222260305

[6] J. Hannan, "Approximation to Bayes risk in repeated plays," *Contributions to the Theory of Games*, vol. 3, pp. 97–139, 1957.

[7] D. P. D. Farias and N. Megiddo, "Combining expert advice in reactive environments," *J. ACM*, vol. 53, no. 5, pp. 762–799, Sep. 2006. [Online]. Available: http://doi.acm.org/10.1145/1183907.1183911

[8] Apple Support, "iOS: Understanding iBeacon," https://support.apple.com/en-gb/HT202880, Feb. 2015.

[9] R. Rodrigues, D. Barnard-Wills, D. Wright, P. De Hert, V. Papakonstantinou, L. Beslay, E. JRC-IPSC, N. Dubois, and E. JUST, "EU privacy seals project," *Publications Office of the European Union*, 2013.

[10] P. Higgins and L. Tien, "Mobile tracking code of conduct falls short of protecting consumers," https://www.eff.org/deeplinks/2013/10/mobile-tracking-code-conduct-falls\-short-protecting-consumers, October 2013.

[11] L. Demir, M. Cunche, and C. Lauradoux, "Analysing the privacy policies of Wi-Fi trackers," in *Workshop on Physical Analytics*, Bretton Woods, USA, Jun. 2014. [Online]. Available: http://hal.inria.fr/hal-00983363

[12] , "Apple - privacy built in," https://www.apple.com/privacy/privacy-built-in/.

[13] M. Gruteser and D. Grunwald, "Enhancing location privacy in wireless LAN through disposable interface identifiers: A quantitative analysis," *Mob. Netw. Appl.*, vol. 10, no. 3, pp. 315–325, Jun. 2005.

[14] T. Jiang, H. J. Wang, and Y.-C. Hu, "Preserving location privacy in wireless LANs," in *Proceedings of MobiSys '07*, 2007, pp. 246–257. [Online]. Available: http://doi.acm.org/10.1145/1247660.1247689

[15] M. Li, K. Sampigethaya, L. Huang, and R. Poovendran, "Swing & swap: User-centric approaches towards maximizing location privacy," in *Proceedings of WPES '06*, 2006, pp. 19–28. [Online]. Available: http://doi.acm.org/10.1145/1179601.1179605

[16] C. Riederer, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and P. Rodriguez, "For sale : Your data: By : You," in *Proceedings of HotNets-X*. New York, NY, USA: ACM, 2011, pp. 13:1–13:6. [Online]. Available: http://doi.acm.org/10.1145/2070562.2070575

[17] A. Ghosh and A. Roth, "Selling privacy at auction," in *Proceedings of EC '11*, 2011, pp. 199–208. [Online]. Available: http://doi.acm.org/10.1145/1993574.1993605

[18] R. Shokri, "Privacy games: Optimal user-centric data obfuscation," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 2, pp. 1–17, 2015.

[19] P. Kumaraguru and L. F. Cranor, "Privacy Indexes: A Survey of Westin's Studies," Carnegie Mellon University, Institute for Software Research International, Tech. Rep., 12 2005.

[20] V. Rastogi and S. Nath, "Differentially private aggregation of distributed time-series with transformation and encryption," in *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '10. New York, NY, USA: ACM, 2010, pp. 735–746. [Online]. Available: http://doi.acm.org/10.1145/1807167.1807247

[21] O. Abul, F. Bonchi, and M. Nanni, "Never walk alone: Uncertainty for anonymity in moving objects databases," in *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*, ser. ICDE '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 376–385. [Online]. Available: http://dx.doi.org/10.1109/ICDE.2008.4497446

[22] M. Terrovitis and N. Mamoulis, "Privacy preservation in the publication of trajectories," in *Proceedings of the The Ninth International Conference on Mobile Data Management*, ser. MDM '08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 65–72. [Online]. Available: http://dx.doi.org/10.1109/MDM.2008.29

[23] R. Chen, G. Acs, and C. Castelluccia, "Differentially private sequential data publication via variable-length n-grams," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, ser. CCS '12. New York, NY, USA: ACM, 2012, pp. 638–649. [Online]. Available: http://doi.acm.org/10.1145/2382196.2382263

[24] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, ser. CCS '13. New York, NY, USA: ACM, 2013, pp. 901–914. [Online]. Available: http://doi.acm.org/10.1145/2508859.2516735

[25] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "A predictive differentially-private mechanism for mobility traces," in *Privacy Enhancing Technologies*. Springer, 2014, pp. 21–41.

[26] G. Barthe, B. Köpf, F. Olmedo, and S. Zanella Béguelin, "Probabilistic relational reasoning for differential privacy," in *Proceedings of the 39th Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, ser. POPL '12. New York, NY, USA: ACM, 2012, pp. 97–110. [Online]. Available: http://doi.acm.org/10.1145/2103656.2103670

[27] J. Reed and B. C. Pierce, "Distance makes the types grow stronger: A calculus for differential privacy," in *Proceedings of the 15th ACM SIGPLAN International Conference on Functional Programming*, ser. ICFP '10. New York, NY, USA: ACM, 2010, pp. 157–168. [Online]. Available: http://doi.acm.org/10.1145/1863543.1863568

[28] K. Chatzikokolakis, M. Andrés, N. Bordenabe, and C. Palamidessi, "Broadening the scope of differential

privacy using metrics," in *Privacy Enhancing Technologies*, ser. Lecture Notes in Computer Science, E. De Cristofaro and M. Wright, Eds. Springer Berlin Heidelberg, 2013, vol. 7981, pp. 82–102. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-39077-7_5

[29] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the Third Conference on Theory of Cryptography*, ser. TCC'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 265–284. [Online]. Available: http://dx.doi.org/10.1007/11681878_14

[30] G. Miklau and D. Suciu, "A formal analysis of information disclosure in data exchange," in *Proceedings of SIGMOD '04*, 2004, pp. 575–586. [Online]. Available: http://doi.acm.org/10.1145/1007568.1007633

[31] C. Goodwin, "A conceptualization of motives to seek privacy for nondeviant consumption," *Journal of Consumer Psychology*, vol. 1, no. 3, pp. 261 – 284, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1057740808800393

[32] B. Huberman, E. Adar, and L. Fine, "Valuating privacy," *Security Privacy, IEEE*, vol. 3, no. 5, pp. 22–25.

[33] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S001999586590241X

[34] J. Demko-Rihter and I. t. Halle, "Revival of high street retailing – the added value of shopping apps," *The AMFITEATRU ECONOMIC journal*, vol. 17, no. 39, 2015. [Online]. Available: http://EconPapers.repec.org/RePEc:aes:amfeco:v:39:y:2015:i:17:p:632

[35] H. Jang, I. Ko, and J. Kim, "The effect of group-buy social commerce and coupon on satisfaction and continuance intention – focusing on the expectation confirmation model (ecm)," in *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, Jan 2013, pp. 2938–2948.

[36] T. Kowatsch and W. Maass, "In-store consumer behavior: How mobile recommendation agents influence usage intentions, product purchases, and store preferences," *Computers in Human Behavior*, vol. 26, no. 4, pp. 697 – 704, 2010, emerging and Scripted Roles in Computer-supported Collaborative Learning. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563210000087

[37] I. M. Dinner, H. J. Van Heerde, and S. Neslin, "Creating Customer Engagement Via Mobile Apps:How App Usage Drives Purchase Behavior," *Social Science Research Network Working Paper Series*, Oct. [Online]. Available: http://ssrn.com/abstract=2669817

[38] J.-Y. M. Kang, J. M. Mun, and K. K. Johnson, "In-store mobile usage: Downloading and usage intention toward mobile location-based retail apps," *Computers in Human Behavior*, vol. 46, pp. 210 – 217, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0747563215000242

[39] Sales Force, "2014 Mobile Behavior Report," https://www.exacttarget.com/sites/exacttarget/files/deliverables/etmc-2014mobilebehaviorreport.pdf, Feb. 2014.

[40] K. L. S. Ohri, "The New Digital Divide: Retailers, shoppers, and the digital influence factor," http://www2.deloitte.com/content/dam/Deloitte/us/Documents/consumer-business/us-rd-thenewdigitaldivide-041814.pdf, 2014.

[41] D. Kosir, "Mobile apps vs. mobile web: What retailers need to know," http://clearbridgemobile.com/mobile-apps-vs-mobile-web-what-retailers-need-to-know/, Aug. 2015.

[42] R. Libfrand, "Retail Mobile App Users Visit Brick-and-Mortars More Often," http://blog.compariscope.com/retail-mobile-apps-12x-the-number-of-in-store-visits, Jan. 2016.

[43] C. Boyle, "Mobile Messaging Trends—Tapping into SMS, Mobile Email and Push," http://www.slideshare.net/eMarketerInc/emarketer-webinar-mobile-messaging-trendstapping-into-sms-mobile-email-and-push-25068768, Aug. 2013.

[44] C. Shepard, A. Rahmati, C. Tossell, L. Zhong, and P. Kortum, "Livelab: Measuring wireless networks and smartphone users in the field," *SIGMETRICS Perform. Eval. Rev.*, vol. 38, no. 3, pp. 15–20, Jan. 2011. [Online]. Available: http://doi.acm.org/10.1145/1925019.1925023

[45] A. Nandugudi, A. Maiti, T. Ki, F. Bulut, M. Demirbas, T. Kosar, C. Qiao, S. Y. Ko, and G. Challen, "Phonelab: A large programmable smartphone testbed," in *Proceedings of SENSEMINE'13*, 2013, pp. 4:1–4:6. [Online]. Available: http://doi.acm.org/10.1145/2536714.2536718

[46] P. Yin, P. Luo, W.-C. Lee, and M. Wang, "App recommendation: A contest between satisfaction and temptation," in *Proceedings of WSDM '13*. New York, NY, USA: ACM, 2013, pp. 395–404. [Online]. Available: http://doi.acm.org/10.1145/2433396.2433446

[47] aestetix and C. Petro, "CRAWDAD data set hope/amd (v. 2008-08-07)," Downloaded from http://crawdad.org/hope/amd/, Aug. 2008.

[48] T. Goodspeed and N. Filardo, "CRAWDAD data set hope/nh_amd (v. 2010-07-18)," Downloaded from http://crawdad.org/hope/nh_amd/, Jul. 2010.

[49] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong, "CRAWDAD data set ncsu/mobilitymodels (v. 2009-07-23)," Downloaded from http://crawdad.org/ncsu/mobilitymodels/, Jul. 2009.

[50] J. Little and B. O'Brien, "A technical review of cisco's wi-fi-based location analytics," http://www.cisco.com/c/en/us/products/collateral/wireless/mobility-services-engine/white_paper_c11-728970.pdf, July 2013.

[51] Derek Top, "Indoor Location Firm Nomi Faces Layoffs; Privacy Concerns To Blame?" http://t.co/e7Gp7mU1Sz, Aug. 2014.

Appendix A

In the following, the user traverses a path of a length m in an area containing $|Z|$ zones. PR-LBS releases a path, $path_{obs}$, to the service provider.

D.P. Mechanism Variant Privacy Guarantees

For the mechanism to achieve (ϵ, d_m) differential privacy, it must satisfy:

$$P(path_{obs}|path) \leq e^{\epsilon} P(path_{obs}|path') \tag{9}$$

such that $d(path, path') \leq d_m$ and $P(path_{obs}|path_{tr})$ is the probability of observing $path_{obs}$ given the user traversed $path_{tr}$. $P(path_{obs}|path)$ is given by:

$$P(path_{obs}|path) = q^{m-d}(1-q)^d \frac{1}{(|Z|-1)^d} \tag{10}$$

such that $d = d(path_{obs}|path)$ is the edit distance between the two paths.

Then, we have:

$$\frac{P(path_{obs}|path)}{P(path_{obs}|path')} \leq e^{\epsilon} \tag{11}$$

$$\frac{q^{m-d}(1-q)^d \frac{1}{(|Z|-1)^d}}{q^{m-d'}(1-q)^{d'} \frac{1}{(|Z|-1)^{d'}}} \leq e^{\epsilon} \tag{12}$$

$$\left(\frac{1-q}{q(|Z|-1)}\right)^{d-d'} \leq e^{\epsilon} \tag{13}$$

When $d(path, path') \leq d_m$, then $d - d' \leq d_m$, because $d > 0$ and $d' > 0$ then $d(path_{obs}, path) - (path_{obs}, path') < |d(path_{obs}, path) - (path_{obs}, path')| < d(path, path')$ by using the reverse triangle inequality for the metric spaces (the edit distance is a metric). Then we have:

$$\left(\frac{1-q}{q(|Z|-1)}\right)^{d-d'} \leq \left(\frac{1-q}{q(|Z|-1)}\right)^{d_m} \leq e^{\epsilon} \tag{14}$$

Finally, this mechanism will achieve (ϵ, d_m) differential privacy for:

$$q \leq \frac{e^{\epsilon/d_m}}{|Z| - 1 + e^{\epsilon/d_m}} \tag{15}$$

D.P. Mechanism Privacy Guarantees

After releasing m zones, the D.P. mechanism satisfies the expression in Eq. (9) for any value of d_m. The probability of observing a path given some other traversed path is given by:

$$P(path_{obs}|path) = \prod (q_i/|Z|_i)^{n_i} \tag{16}$$

Where $|Z|_i$ represents the number of zones at a distance i from user's zones and $q_i/|Z|_i$ the probability to choose a zone from those at a distance i from the visited zone. n_i represents the number of released zones that are a distance i from the actual zones. For two paths at a distance d_m from each other, we need to satisfy the following:

$$\frac{\prod (q_i/|Z|_i)^{n_i}}{\prod (q_j/|Z|_j)^{n_j}} \leq e^{\epsilon};\ i, j \leq d_m \tag{17}$$

The expression of Eq. (17) will assume its maximum value when the observed path is the user's actual path and $path'$ is a path at a distance d_m. In such a case we have (for a path of length m):

$$\frac{q_0^{d_m}}{(\alpha^{d_m} q_0/|Z|_{d_m})^{d_m}} \leq e^{\epsilon} \tag{18}$$

Where $|Z|_{d_m}$ is the maximum number of zones at a distance of d_m from any of the zones of the released path.

We can then derive an lower bound for α and upper bound for q_0 as:

$$\alpha \geq \frac{|Z|_{d_m}}{e^{\epsilon/m}};\ q_0 \leq \frac{1}{\sum \alpha^i}$$

A.S. Mechanism Anonymity Set

The size of the anonymity set is a random variable, S, depends on the length of the released path. Let R be a random variable that represents the length of the released path and k represent the possible value of the traversed path length such that $k \leq m$. Our objective is to compute the expected size of the anonymity set $E(S)$.

$$E(S) = \sum_{k=0}^{m} \sum_{r=0}^{k} E(S|R=r).P(R=r)$$

$$= \sum_{k=0}^{m} \sum_{r=0}^{k} \binom{k}{r} \binom{|Z|-r}{k-r} (k-r)! \binom{k}{r} q^r (1-q)^{k-r}$$

$$= \sum_{k=0}^{m} \sum_{r=0}^{k} \binom{k}{r}^2 q^r (1-q)^{k-r} \frac{(|Z|-r)!}{(|Z|-k)!}$$

Per-Path Information Disclosure

When the user moves to a new zone z_l, PR-LBS estimates the information leak if the SP is to observe that the user visited z_l, with total observed path being $pa_l = \langle pa_{l-1}, z_l \rangle$.

Let $visit(z)$ be the event that the SP observed user visited the zone z. $P(pa_l|visit(z_l))$, then, refers to the

probability distribution of the user visiting the current path pa_l of length l after observing the visit to zone z_l. By definition, a new observation will necessarily change $P(p_l)$ (the SP's existing belief about the user's mobility); the amount of change in this PDF is what we refer to as the information leak. The amount of leaked information can be defined as:

$$lk(p_l, z) = \sup_{p_l \in P_l} \frac{P(p_l|visit(z)) - P(p_l)}{P(p_l)}. \quad (19)$$

Since we focus on the positive information disclosure, the only path that will experience a positive improvement in the amount of information is pa_l, the path the user is currently visiting. This reduces Eq. (19) to:

$$lk(pa_l, z) = \frac{P(pa_l|visit(z)) - P(pa_l)}{P(pa_l)}. \quad (20)$$

If the user has visited the area N times (number of sessions), out of which s/he traversed the path pa_l for $n[pa_l]$ times, then $P(pa_l) = \frac{n[pa_l]}{N}$. When observing a new visit, the probability will be $P(pa_l|visit(z)) = \frac{n[pa_l]+1}{N+1}$. The information leak will then be:

$$lk(pa_l, z) = \frac{1-a}{a(N+1)}, a = \frac{n[pa_l]}{N}.$$

Appendix B

Suppose the user visits an area comprised of four zones: "A", "B", "C", and "D". After 10 sessions, the user's mobility model is as follows:

1. **Paths of length=4:** P($A-B-C-D$)= 2/10, P($A-C-B-D$)=3/10, P($A-D-B-A$)=1/10, P($B-D-C-B$)=1/10, P($B-D-C-A$)=1/10, and P(ϕ)=2/10.
2. **Paths of length=3:** P($A-B-C$)= 2/10, P($A-C-B$)=3/10, P($A-D-B$)=1/10, P($B-D-C$)=2/10, and P(ϕ)=2/10.
3. **Paths of length=3:** P($A-B$)= 2/10, P($A-C$)=3/10, P($A-D$)=1/10, P($B-D$)=2/10, and P(ϕ)=2/10.
4. **Paths of length=3:** P(A)= 6/10, P(B)=2/10, and P(ϕ)=2/10.

During the 11^{th} session, the user traverses the path $B-D-C-A$.

1. First visited zone is B; the path will only comprise B at this point. At the beginning of the visit, the SP expects the user (based on previous mobility) to visit either A or B. Since the user visited B, there is an information leakage.
2. Second visited zone is D; the new path will be $B-D$. This path leaks some information because there are multiple expected paths of length 2. But the visited zone leaks no information according to our criterion. The only expected visited zone after B is D. The user conformed to the SP's expectations and leaked no information. It is worth noting that the information leak of the path thus far is equal to information leak from the first visited zone, which is B. The same applies for the third visited zone C.
3. Last visited zone is A'; the path comprises $B-D-C-A$. The newly visited zone leaks some information since there are two possibilities after traversing $B-D-C$, either A or B. By visiting D, the user offered the SP new information that resulted in a shift of its belief about the user mobility.

Table 6. Privacy cost of visited path.

Zone (z)	Path (p_l)	$P(p_l)$	$P(p_l\|z)$	$lk(p_l, z)$	$leak(z)$
B	B	2/10	3/11	$log_2(\frac{15}{11})$	$log_2(\frac{15}{11})$
D	$B-D$	2/10	3/11	$log_2(\frac{15}{11})$	0
C	$B-D-C$	2/10	3/11	$log_2(\frac{15}{11})$	0
A	$B-D-C-A$	1/10	2/11	$log_2(\frac{20}{11})$	$log_2(\frac{4}{3})$

Appendix C

Figs 13 and 14 show that PR-LBS adapts sharing to the service level. When the server is more rewarding, as in Fig. 13, PR-LBS shares more of the user's mobility. All the utility metrics for the neutral and service-oriented users are close to 1. While those for the privacy oriented users are lower so that PR-LBS protects their privacy. On the other hand, Fig. 14 shows the utility metrics for a low-rewarding server. It is evident that the utility metrics drop considerably when compared to the high-rewarding service provider.

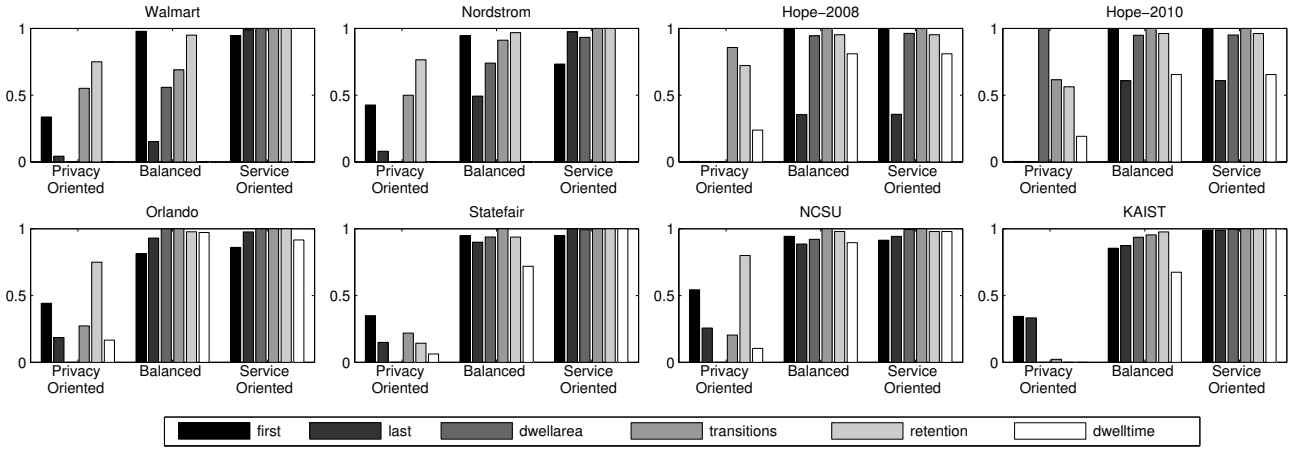


Fig. 13. Utility metrics for high service level. Some metrics in Walmart and Nordstrom datasets are not available.

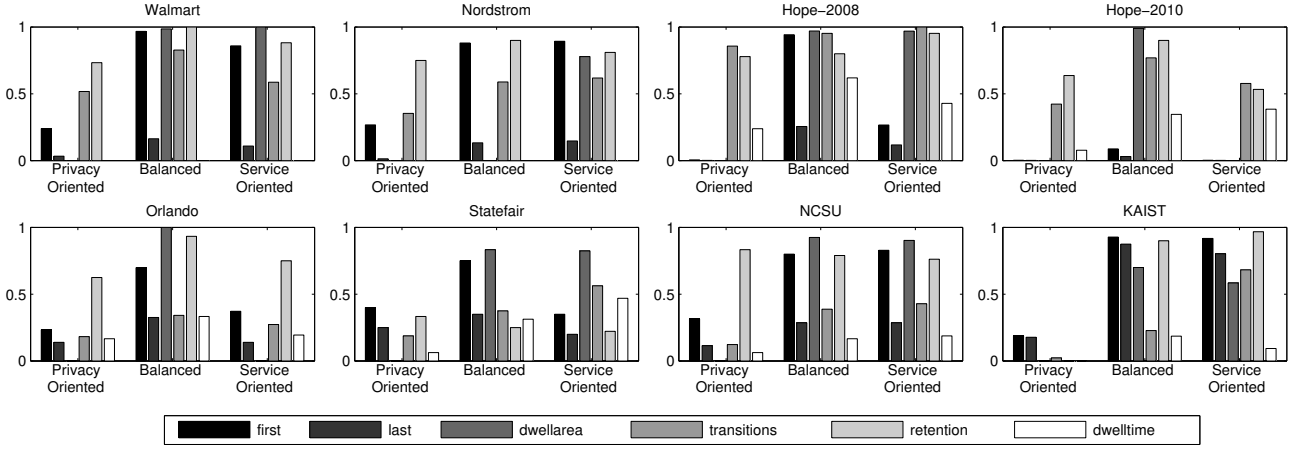


Fig. 14. Utility metrics for low service level. Some metrics in Walmart and Nordstrom datasets are not available.