Protecting Privacy of User Information in Continuous Location-Based Services

Claudio A. Ardagna, Giovanni Livraga, Pierangela Samarati
Dipartimento di Informatica
Università degli Studi di Milano
Crema, Italy
email: *firstname.lastname*@unimi.it

***Abstract*—The widespread diffusion of mobile devices integrating location capabilities makes the location of users yet another type of sensitive information used by service providers in the provision of accurate and personalized services (location-based services – LBSs). A major problem in this context is that the privacy of users is increasingly at risk, calling for solutions balancing the benefits provided by LBSs and the privacy guarantees. In this paper, we study a novel privacy problem related to inferences of sensitive information caused by the release of consecutive positions to LBS providers. We provide an approach based on Markov chains that allows the user to continuously release her location information in a privacy-preserving way. We then define an approach to counteract different inference channels, addressing users' preferences in terms of both privacy requirements and quality of service.**

***Index Terms*—Continuous LBSs, Inference, Location Privacy, Markov Chain**

I. Introduction

As mobile devices gain increasing popularity, information about the physical location of users plays every day a more central role in the provisioning of accurate services to users. We all handle devices (e.g., cellular phones) enriched with capabilities to determine our position and communicate it to service providers. At the same time, many service providers are implementing *Location-Based Services* (LBSs) whose provision is based upon the knowledge of users' position. As a matter of fact, enriching services with location-based functionalities can bring important benefits to the quality of the provided applications, as users can enjoy the benefits conveyed by an accurate and personalized service. As an example, a LBS providing run-time road traffic information can release the traffic status only in the proximity of user's position, rather than for the entire road network of a county. However, such convenience comes at the price of an increasing privacy risk for users releasing their private data to LBSs. In this context, there is a growing demand for solutions that ensure a proper protection of users' location privacy, as witnessed by 55% of LBS users showing concerns about the protection of their privacy when interacting in a mobile environment [1].

Responding to the growing demand for privacy protection in LBSs, in recent years the research community has addressed the problem and proposed several solutions for guaranteeing proper protection to users' identity, location, and personal information in the framework of LBSs. Such large body of research mainly focused on the protection of users' privacy when a single location is released to the service provider and proposed two classes of techniques: anonymity-based techniques [2], which aim at protecting the link between users' identity and their sensitive information, and obfuscation-based techniques [3], which aim at protecting the sensitive information of the users by degrading its accuracy.

Recently, a new line of work has focused on addressing the problem of path privacy [1] with the main goal of anonymizing the footsteps released by the users in their communications with the LBS providers. Such techniques aim at producing a path which is shared by multiple users making them and their interactions with LBS providers indistinguishable. In this paper, we address a novel privacy problem that consists in limiting the amount of inference that can be drawn by an observer accessing the path followed by users. Differently from existing approaches [1], we assume that the paths released by users are linked to their identity. The contribution of our work is therefore threefold. We first model our problem by defining three classes of inferences that exploit private information of the users: inference on sensitive positions that a user has visited, inference on sensitive paths that a user has traversed, and inference on unusual paths taken by the user. We then propose a characterization of the user's behavior as a Markov chain, and provide a simple and intuitive mechanism allowing users to specify preferences on the release of their data. We finally propose a solution, based on user's behavior and preferences, which obfuscates the real path of the users to limit inferences on sensitive information not intended for disclosure (e.g., the fact that a user has driven for the first time to a clinic for rare diseases).

The remainder of this paper is organized as follows. Section II presents our motivations and reference scenario. Section III describes a summary of our approach. Section IV illustrates how users can specify privacy preferences on different location information and quality of service requirements. Section V describes the modeling of our system. Section VI presents our privacy approach to limit inference attacks on location information. Section VII discusses related work. Finally, Section VIII gives our final remarks.

II. Reference Scenario and Motivations

Our reference scenario is a mobile and service-based infrastructure involving mobile users and LBS providers. Mobile users carry mobile devices supporting both GSM/3G and GPS protocols for communication. LBS providers offer services requiring continuous sampling of users' positions (e.g., LBSs for social networks, tracking services, friend finder). Mobile users, moving around a bi-dimensional space, release their positions to LBS providers to enjoy location-based services. Location information released by a user is univocally associated with her, making different releases by the same user linkable. In the remainder of this paper, we will refer to the following running example.

Example 2.1 (Running example): Bob is a user who lives and works in the city of Milan. Bob is subscribed to a continuous LBSs and follows repetitive paths to move between home, work, preferred supermarket and gym. Among his repetitive movements, Bob is used to reach Milan downtown after work for a drink. This path, though recurrent, is sensitive for him, since he wants to hide this information to external observers (*sensitive movement*). One morning, while at work, Bob receives a call informing him that his mother has been transported to the emergency room of a clinic specialized in cardiovascular problems. Bob rushes to the emergency room, which is in the path between his home and the place where he works, to meet the head physician taking care of his mother. Bob considers being at the emergency room a sensitive information (*sensitive position*). Also, the week after the heart attack, Bob visits a different clinic to get a second medical advice on his mother's disease. To reach this clinic, Bob moves along an unusual path, which can be exploited for inference by a LBS provider (*unusual path*).

The above running example identifies three specific inference channels that exploit different information on the users' positions: sensitive positions, sensitive movements, and unusual paths. In the following, we better detail how these three kinds of information may cause inference channels.

- *Sensitive positions.* The knowledge that a user has visited a certain place can disclose sensitive information. For instance, the fact that Bob is visiting the emergency room of a clinic specialized in cardiovascular problems can disclose the information that he or one of his relatives may suffer from a cardiovascular disease. The definition of sensitive positions is user dependent, since different users may have different perceptions. As an example, while Bob, aiming to keep private health information, may consider being in a hospital a sensitive position, Alice may be willing to disclose the fact that she is at the hospital and consider sensitive the information of being in religious places.
- *Sensitive movements.* Some paths may be sensitive since they leak private information. For instance, the fact that Bob is walking in Milan downtown every day after work can disclose information about his lifestyle. Recalling that location data released by a user to the LBS provider are linked to the user's identity, the LBS provider can observe the footprints of each user on the road network. Inferences exploiting sensitive paths are due to the observation that a user followed a path on the road network which is maybe sensitive, and possibly recurrent. We note that a sensitive path can be composed of an arbitrary number of movements. In this paper, without loss of generality, we consider sensitive paths composed of a single movement, as the definition of a sensitive path in terms of a "sensitive movement" implicitly captures the more general definition of a sensitive path composed of multiple consecutive movements. Like for sensitive positions, the sensitivity associated with a specific movement depends on the individual user's perception.
- *Unusual paths.* Users' paths are usually repetitive and strongly depend on users' profile (e.g., the place where they work/live). We assume these paths, called *usual paths*, to be known by the observer. Inferences exploiting *unusual paths* are due to the observation of *deviations* in the path followed by a user with respect to usual paths. For instance, in our example, Bob takes an unusual path to visit a clinic specialized in cardiovascular diseases for a second advice. An observer noticing this deviation may infer sensitive health information related to Bob. We note that what causes this inference is the deviation from the usual path, rather than the path itself. Differently from sensitive positions and movements, the sensitivity of an unusual path is intrinsic to the path itself, and does not depend on the perception of the individual users.

The motivation of our work is to provide users with an easy-to-use and intuitive solution to limit the possible inferences that an observer can draw from the knowledge of users' positions. Our goal is therefore to design an approach for protecting the privacy of the users against the three inference channels introduced in this section.

III. Sketch of our Approach

We consider a privacy architecture where all communications between mobile users and LBS providers are mediated and filtered by a *trusted privacy middleware* so to block inference channels. The trusted middleware evaluates the location information released by the user, assesses the risk of inference, and possibly obfuscates the path of the user before releasing it to a LBS provider. In other words, as shown in Figure 1, the privacy middleware enforces users' preferences and obfuscates users' paths, to create cover stories that are released to the LBS provider for protecting users against the inference channels based on sensitive positions, sensitive movements, and unusual paths. The cover story must satisfy three basic properties: *i)* being safe w.r.t. inference channels; *ii)* being realistic, so that an observer cannot detect the cover stories among the released location information; and *iii)* minimizing the distortion w.r.t. the real information, so to preserve as much as possible the quality of the provided service. We note that given the increasing computational resources of mobile devices, the

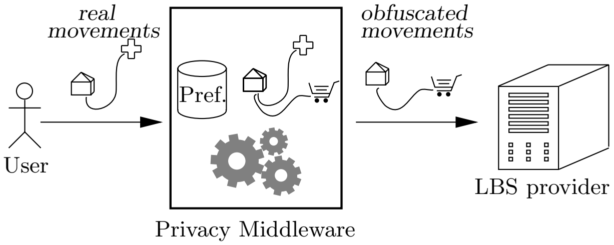


Fig. 1: Privacy Architecture

privacy middleware can be installed on users' device with no need of a trusted third party.

To address the above properties, our solution assumes a delay in the release of location information to the LBS provider (so to enable possible inspection of cover stories). Then, the privacy middleware releases the obfuscated location at time *t* based on the observed path at time *t+k*. This is a feasible assumption in many real-world scenarios (e.g., having a tracking service or a friend finder with a delay of few seconds does not degrade the quality of the service), and allows the privacy middleware to evaluate the risk of inference and to generate the cover story using additional locations. In fact, based on the introduced delay, the middleware evaluates the movement trends of the users at time *t+k* to adjust the cover story at time *t* and address both privacy and service quality preferences.

The activities of the privacy middleware are driven by privacy preferences. These preferences are specified by each user and identify those sensitive positions and/or movements which may cause inference on private information. Our solution uses these preferences to counteract the inference channels in Section II. Intuitively, a different cover story will be defined by the privacy middleware depending on the inference channels as follows: *i) sensitive position*, a cover story will be produced such that it does not end in a sensitive position; *ii) sensitive movements*, a cover story will be produced such that it does not contain sensitive movements; *iii) unusual path*, a cover story will be produced such that the unusual path is mapped to a usual one.

IV. System Modeling and Users' Preferences

We consider a mobile user u moving in a bi-dimensional space $\mathcal{S}$ constrained by a road network. We model the road network as a graph $G(V,E)$ with vertices in V representing road intersections (e.g., roundabouts, traffic lights) and points of interest (e.g., emergency room, shopping center, gym) and edges in E representing roads. Figure 2 shows an example of graph for the running example in Example 2.1. In the remainder of the paper, we denote vertices that refer to road intersections and points of interest with their initial letter (e.g., we refer to the emergency room with $\mathtt{e}$).

Based on the graph $G(V,E)$, we provide a mechanism that allows each user to express her privacy preferences in a simple and intuitive way. In our framing of the problem, each user maintains a set of preferences (i.e., sensitivity levels) for different positions and/or movements, which indirectly specify the sensitive information that she aims to safeguard.

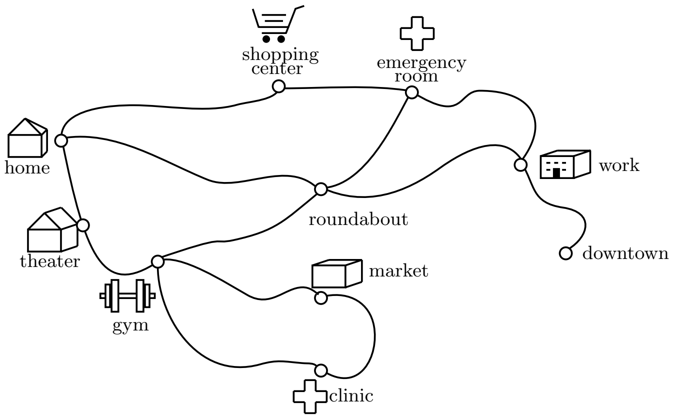


Fig. 2: Graph representation of the network in Example 2.1

We allow users to define two different types of preferences: *privacy preferences* that specify *positions* and *movements* that a user perceives as sensitive, and *service quality preferences* that specify a minimum quality level that must be guaranteed by the location-based service. We assume unusual paths to be sensitive by default.

In the following we illustrate how the above preferences can be specified by the users on $G(V,E)$.

Privacy preferences. Users express privacy preferences as *sensitivity labels* that they can associate with both vertices and edges of the graph $G(V,E)$. Sensitivity labels applied by user u to vertices (edges, resp.) of $G(V,E)$ capture how much u considers sensitive the fact that she has visited the associated location (made the associated movement, resp.). In this paper, for simplicity but without loss of generality, we consider a set of labels composed by two elements, 0 and 1, representing a non-sensitive and sensitive location information, respectively. The set of possible sensitivity labels could be extended to *any* set, composed of an arbitrary number of labels, provided the existence of a (partial) order relationship over them.

Users can specify the sensitivity of the vertices and edges in $G(V,E)$ via a labeling function λ:$V \cup E \rightarrow \{0,1\}$, to counteract inference attacks. In particular, users specify privacy preferences on vertices of $G(V,E)$ to represent sensitive positions. The sensitivity label states whether the user considers a sensitive information the fact that she stopped in the position represented by v ($\lambda(v)$=1) or not ($\lambda(v)$=0). Sensitive positions form a subset $V_{sens} \subseteq V$:$\forall v \in V_{sens}$,$\lambda(v)$=1, which includes *points of interest* that a user considers sensitive. For instance, with reference to our running example, V_{sens}=$\{\mathtt{e}\}$. In fact, Bob considers the emergency room of the clinic for cardiovascular diseases in vertex $\mathtt{e}$ as the only sensitive position and, accordingly, he specifies $\lambda(\mathtt{e})$=1. All other positions are not sensitive (e.g., $\lambda(\mathtt{h})$=0, since he does not consider his house a sensitive position).

Similarly, users specify privacy preferences on edges of $G(V,E)$ to represent sensitive movements. The sensitivity label states whether the user considers a sensitive information the fact that she traversed the road represented by (v_i, v_j) $(\lambda((v_i, v_j))=1)$ or not $(\lambda((v_i, v_j))=0)$. Sensitive movements form a subset $E_{sens} \subseteq E$:$\forall (v_i, v_j) \in E_{sens}, \lambda((v_i, v_j))=1$, which includes *movements* that a user considers sensitive. For instance, with reference to our running example, $E_{sens}=\{(\mathtt{w}, \mathtt{d})\}$. In fact, Bob considers his walking from work (vertex w) to the city downtown (vertex d) for a drink the only sensitive movement, and specifies $\lambda((\mathtt{w}, \mathtt{d}))=1$.

We note that a default preference policy can be defined labeling all vertices and edges as sensitive (non-sensitive, resp.). Users can also specify preferences at different levels of granularity. For instance, in Example 2.1, Bob can either specify each clinic as sensitive, or specify CLINICS as a sensitive category. In the latter case, we assume a pre-processing phase translating preferences on sensitive categories into preferences on specific positions on the space $\mathcal{S}$. For instance, if Bob specifies CLINICS as a sensitive category, this preference can be used to protect movements of Bob who stops at either e or c in Figure 2.

Service quality preferences. As previously mentioned, we counteract inferences by releasing to the LBS provider cover stories that obfuscate the real position (path, resp.) of a user with an alternative one. Intuitively, the more the distance between the obfuscated position (path, resp.) and the real one, the higher the privacy protection but the poorer the received service quality. The goal of service quality preferences is therefore to impose an upper bound to the level of obfuscation applied to the real position (path, resp.), to balance privacy protection and service quality. As an example, suppose that Bob is moving downtown Milan and his cover story is releasing a position in the suburbs of Milan. If, on one side, the cover story gives high privacy, on the other side, the quality of the service is highly compromised. We capture this need by allowing a user to define the *maximum distance* δ that can exist between her real position and the corresponding one in the cover story. For instance, with reference to our running example, Bob can choose δ=200, meaning that the maximum distance between his real position and the position disclosed by the cover story cannot be higher than 200 meters.

The above preferences form the *inference policy* of a user, which is formally defined as follows.

Definition 4.1 (Inference Policy): Let u be a user and $G(V,E)$ the graph of the road network in $\mathcal{S}$. The *inference policy* $\mathcal{P}_u$ of u is a triple $(V_{sens}, E_{sens}, \delta)$ such that: $V_{sens} \subseteq V$, $E_{sens} \subseteq E$, and $\delta \in \mathbb{R}$.

Example 4.2: Consider our running example in Example 2.1 and the graph in Figure 2. The inference policy of Bob is represented by the triple $\mathcal{P}_{Bob} = (\{\mathtt{e}\}, \{(\mathtt{w}, \mathtt{d})\}, 200)$, stating that: *i)* he has a privacy preference on sensitive positions ($V_{sens}=\{\mathtt{e}\}$); *ii)* he has a privacy preference on sensitive movements ($E_{sens}=\{(\mathtt{w}, \mathtt{d})\}$); and *iii)* he has a service quality preference on the maximum distance (δ=200).

V. MODELING USERS' MOVEMENTS

To properly protect the privacy of a user against the inference channels described in Section II, we need to capture the moving behavior of the user, that is, the usual paths followed by her when moving in $\mathcal{S}$. The modeling of these paths is necessary here for two reasons: *i)* to produce a realistic cover story, indistinguishable from the real movements of the user; and *ii)* to protect the user against inferences on unusual paths.

In principle, different aspects can be considered to model the moving behavior of a user, such as the context in which the user is moving, the time at which a movement happens, the history of past movements, the reason why the user is moving, or the motion model. In this paper, for the sake of simplicity, we consider the frequency of the movements, that is, the number of times a user u is moving from a vertex v_i to a vertex v_j, to represent her moving behavior.

We observe the movements of a user u on the graph $G(V,E)$ during a *learning step*. In this step, the mobile user simulates a communication with a LBS provider and releases continuous samples of her position to the middleware. The latter enforces user's preferences and, based on the produced cover stories, calculates a *count matrix* specifying the number of times (frequency) the user is moving from a vertex $v_i \in V$ to a vertex $v_j \in V$. The count matrix M associated with V is a square matrix of size $|V| \times |V|$, such that $\forall i=1,\ldots,|V|, \forall j=1,\ldots,|V|$, $c_{ij}=M[i][j]$ represents the number of times u has left position v_i toward position v_j. We note that the reason to generate the count matrix based on the cover story is to avoid inferences based on the real user's movements. For instance, consider Bob's movements in our running example (Example 2.1). Despite considered sensitive, the movement to Milan downtown is usual for Bob, and a count matrix calculated on real movements would include it. After the learning step, when releasing the cover story, the privacy preferences of Bob would cause the sensitive movement towards Milan downtown to be excluded from the released path. This would cause an evident inconsistency between the frequencies in the count matrix and the released cover stories, opening the door to a new inference channel on Bob's privacy preferences. By contrast, if the count matrix is calculated on data enforcing the privacy preferences (see Example 5.1), no inconsistencies will exist between the released cover stories and the count matrix.

Example 5.1: Suppose that in the learning step, for each of the vertices of the space graph in Figure 2, the middleware records 200 movements for Bob. Figure 3(a) shows the count matrix for Bob at the end of the recording phase. Each cell in the matrix shows the number of times Bob has entered a vertex and left toward another vertex. For instance, out of 200 times Bob has left his home in h (row 1), he left 70 times toward the theater in t (column 2), 70 toward the shopping center in s (column 4), 60 toward the roundabout in r (column 5), and never toward g, e, m, c, w and d. Although Bob is moving every day from his workplace to the city downtown, the count matrix does not include this movement. This is due to the fact

	h	t	g	s	r	e	m	c	w	d
h	-	70	-	70	60	-	-	-	-	-
t	100	-	100	-	-	-	-	-	-	-
g	-	60	-	-	40	-	80	20	-	-
s	140	-	-	-	-	60	-	-	-	-
r	20	-	130	-	-	30	-	-	20	-
e	-	-	-	140	40	-	-	-	20	-
m	-	-	160	-	-	-	-	40	-	-
c	-	-	100	-	-	-	100	-	-	-
w	-	-	-	-	80	120	-	-	-	-
d	-	-	-	-	-	-	-	-	-	-

(a)

	h	t	g	s	r	e	m	c	w
h	0.00	0.35	0.00	0.35	0.30	0.00	0.00	0.00	0.00
t	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00
g	0.00	0.30	0.00	0.00	0.20	0.00	0.40	0.10	0.00
s	0.70	0.00	0.00	0.00	0.00	0.30	0.00	0.00	0.00
r	0.10	0.00	0.65	0.00	0.00	0.15	0.00	0.00	0.10
e	0.00	0.00	0.00	0.70	0.20	0.00	0.00	0.00	0.10
m	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.20	0.00
c	0.00	0.00	0.50	0.00	0.00	0.00	0.50	0.00	0.00
w	0.00	0.00	0.00	0.00	0.40	0.60	0.00	0.00	0.00

(b)

Fig. 3: Count matrix (a) and related transition function (b)

that the preferences of Bob are enforced before calculating the matrix frequencies.

The count matrix associated with a user u represent the probability with which u is moving in $G(V,E)$. In fact, the probability that u leaves position $v_i{\in}V$ toward position $v_j{\in}V$, $\forall i,j{=}1,\ldots,|V|$, is trivially computed as: $P(j|i){=}\frac{c_{ij}}{\sum_i c_{ij}}$. For instance, with reference to the count matrix in Figure 3, the probability for Bob of moving between his home in h and the city theater in t is computed as: $c_{\texttt{ht}}/(c_{\texttt{ht}}+c_{\texttt{hs}}+c_{\texttt{hr}}) = 70/200 = 0.35$.

The probabilities $P(j|i), \forall i,j{\in}V$ captured by our count matrix correspond to the transition function of a first-order Markov model [4]. In our modeling, the positions of a user over time are then represented as a set of stochastic variables, that can assume different values in the set V of locations (i.e., *states* in the Markov model). The transition function characterizing a Markov model maps the moving behavior of a user u, that is, the "likelihood" with which u moves in $\mathcal{S}$ and selects the next movement given the current position. The probability between two consecutive points can then be extended to calculate the probability of a path as follows.

Definition 5.2 (Path probability): Given a count matrix M for a user u and a path $\langle v_1,\ldots,v_j\rangle$ of vertices, the *probability* $P(\langle v_1,\ldots,v_j\rangle)$ of $\langle v_1,\ldots,v_j\rangle$ is computed as $\prod_{k=1}^{j-1} P(k{+}1|k)$.

Definition 5.2 computes the probability for u to traverse a path as the joint probability of traversing the corresponding sequence of edges. For instance, the probability for Bob of moving between his home (vertex h), the theater (vertex t), and the gym (vertex g) is computed as: $c_{\texttt{ht}} \times c_{\texttt{tg}}{=}0.35 \times 0.50{=}0.175$.

We note that more than one Markov model can be associated with the same user u. The moving behavior of u, in fact, can change with contextual information (e.g., the road from home to work is very probable during the week, while less probable during the weekends). However, without loss of generality, we assume each user to be associated with a single Markov model, and we note that our model can be straightforwardly extended for capturing the existence of a set thereof.

We define the transition function of the Markov model as a square matrix T, where $T[i][j]{=}P(j|i), \forall i,j{\in}V$. T must satisfy the following properties:

1) $\forall i{=}1,\ldots,|V|$, $\forall j{=}1,\ldots,|V|$, $T[i][j]{\geq}0$, that is, all elements of the matrix T are greater or equal to zero;
2) $\forall i{=}1,\ldots,|V|, \sum_{j=1}^{|V|} T[i][j]{=}1$, that is, the elements on each row of the matrix T sum to 1.

Figure 3(b) illustrates the transition function computed from the count matrix in Figure 3(a) for Bob. We note that to guarantee the above properties, vertices that have never been reached by the user in the count matrix (e.g., vertex d in Figure 3(a)), are not included in T. We consider these areas as *shadow areas*. For the sake of simplicity, in the following of this section, we do not consider shadow areas and we denote with $G(V,E)$ the graph that can be generated starting from the transition function in Figure 3(b). Intuitively, $G(V,E)$ is equal to the graph in Figure 2 without vertex d and edge (w,d).

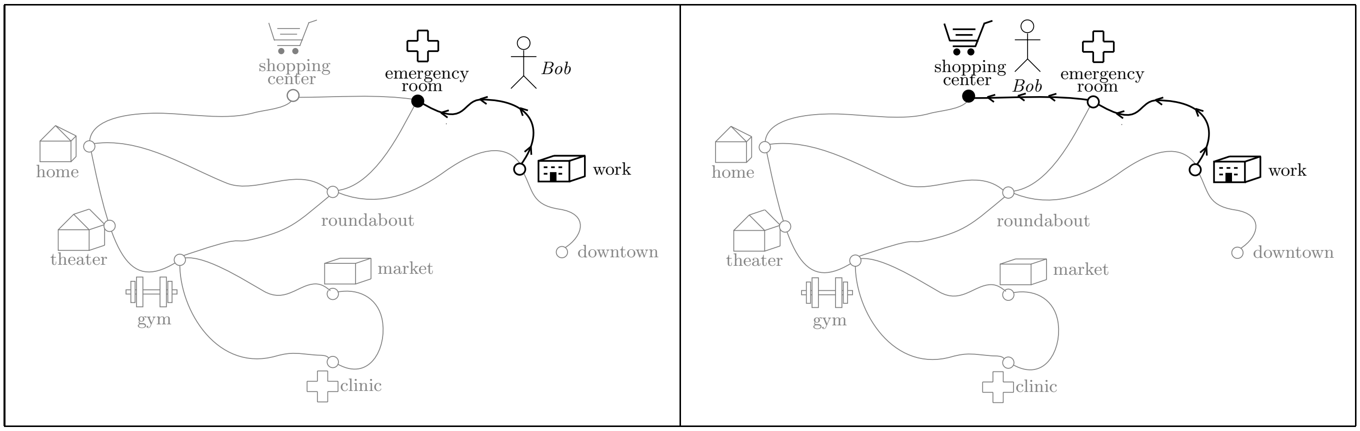
VI. Counteracting Inferences

We describe our solution to protect privacy of mobile users against the three inference channels that can be exploited by a location provider (see Section II). We consider a user u following a path $\langle v_1,\ldots,v_{i-1},v_i\rangle$ in $G(V,E)$ that: *i)* stops in a sensitive position v_i (i.e., such that $\lambda(v_i){=}1$); *ii)* travels along a sensitive movement (v_{i-1},v_i) (i.e., such that $\lambda((v_{i-1},v_i)){=}1$); and *iii)* takes a movement (v_{i-1},v_i) which makes the path unusual.

Inference on sensitive positions. When u stops at v_i and $\lambda(v_i){=}1$, a risk of inference arises and v_i should not be released by the middleware. Our solution builds a cover story in which the path followed by u is obfuscated by the middleware to release a different path $\langle v_1,\ldots,v_j\rangle$, where $i{\neq}j$ and $\lambda(v_j){=}0$. To this aim, we build a set V_{safe} of *safe positions* enforcing the preferences of u, formally defined as follows.

Definition 6.1 (V_{safe}): Let V be the set of vertices modeling road intersections and points of interest in the space $\mathcal{S}$, and v_i the sensitive position at which u stops. A set V_{safe} of safe positions is a subset of V such that $\forall v{\in}V_{safe}$, $\lambda(v){=}0$ and $d(v_i,v) \leq \delta$, with $d(v_i,v)$ the distance between v_i and v, and δ the service quality preference in $\mathcal{P}_u$.

We note that V_{safe} addresses both privacy preferences (i.e., $\lambda(v){=}0$) and service quality preferences (i.e., $d(v_i,v) \leq \delta$). We also note that the set of safe positions, in conjunction with the specification of sensitive positions as categories, allows to counteract inferences on sensitive positions while u is moving in shadow areas. Preferences on sensitive and non-sensitive positions in shadow areas are in fact implicitly specified using categories.



(a) Real path

(b) Released path

Fig. 4: Counteracting inference on sensitive positions

Given $\langle v_1,\ldots,v_i\rangle$ and V_{safe}, the privacy middleware has to select the best position $\hat{v}_j \in V_{safe}$ and to release a path $\langle v_1,\ldots,\hat{v}_j\rangle$ that will simulate the movement of u toward $\hat{v}_j$. The best position $\hat{v}_j$ is the position that is reached with the highest probability (see Definition 5.2). We can then associate a probability $P(\langle v_1,\ldots,v_j\rangle)$ for each $v_j \in V_{safe}$ and select the vertex $\hat{v}_j = v_j$ for which $P(\langle v_1,\ldots,v_j\rangle)$ is maximum. The cover story is then built by the middleware as follows: *i)* if $\hat{v}_j$ is reached after the sensitive position v_i, the middleware releases a set of fake positions toward $\hat{v}_j$; *ii)* if $\hat{v}_j$ is reached before v_i, the middleware has to suppress all positions to be released after $\hat{v}_j$ (thanks to the delay in the release of cover stories discussed in Section III).

In case $|V_{safe}|$=0, no safe positions exist around u. We then propose to alert u and apply an exit strategy choosing a position v such that $\lambda(v)$=1 for u, while $\lambda(v)$=0 for the majority of users. The intuition is that releasing a position safe for the majority of other users can reduce the risk of inference.

Figure 4 shows an example of how inferences on sensitive positions are counteracted. In the figure and in the remainder of the paper, gray vertices and edges are not included in Bob's path, while black vertices (edges, resp.) represent positions where Bob has stopped (movements Bob have done, resp.). Figure 4(a) shows the real path of Bob going from work to the emergency room, and stopping there. Since $\lambda(\mathtt{e})$=1, the middleware computes V_{safe}=$\{\mathtt{s},\mathtt{r}\}$, and creates the cover story with the most probable position according to T (see Figure 3(b)), that is, $\mathtt{s}$. Figure 4(b) illustrates the cover story $\langle \mathtt{w},\mathtt{e},\mathtt{s}\rangle$ released to the LBS provider.

Inference on sensitive movements. When u traverses an edge (v_{i-1},v_i) corresponding to a sensitive movement, a risk of inference arises and (v_{i-1},v_i) should not be released by the middleware. Our solution builds a cover story in which the path followed by u is obfuscated by the middleware to release a different path $\langle v_1,\ldots,v_{i-1},v_j\rangle$, where v_i is substituted by v_j provided that there exists an edge (v_{i-1},v_j). To this aim, we build a set E_{safe} of *safe movements*, formally defined as follows.

Definition 6.2 (E_{safe}): Let E be the set of edges modeling roads in the space $\mathcal{S}$, and (v_{i-1},v_i) the sensitive movement taken by $u_.$ A set E_{safe} of safe movements is a subset of E such that $\forall(v_{i-1},v_j)\in E_{safe}$, $\lambda((v_{i-1},v_j))$=0 and $d(v_i,v_j)\leq\delta$, with $d(v_i,v_j)$ the distance between the final vertices of the real movement (i.e., v_i) and of the safe movement (i.e., v_j), and δ the service quality preference in $\mathcal{P}_u$.

We note that E_{safe} addresses both privacy preferences (i.e., $\lambda((v_{i-1},v_j))$=0) and service quality preferences (i.e., $d(v_i,v_j)\leq\delta$). Given E_{safe}, the privacy middleware has then to select and release the best movement $(\hat{v}_h,\hat{v}_k)\in E_{safe}$ with $\hat{v}_h$=v_{i-1}. We propose a greedy approach that selects the most probable movement $(\hat{v}_h,\hat{v}_k)$, based on Definition 5.2 and T.

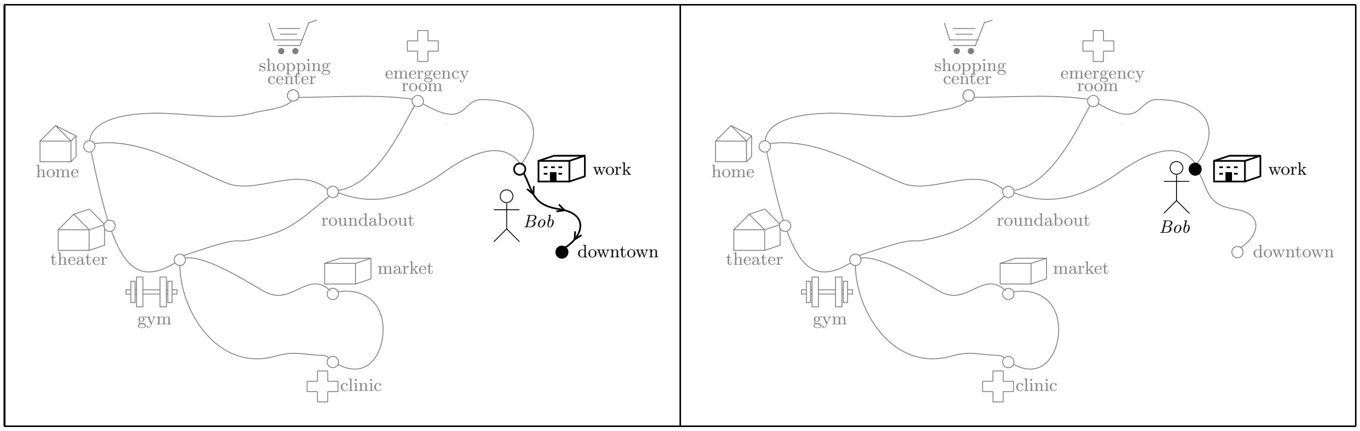
In case $|E_{safe}|$=0, no safe movements exist around u, which end in positions satisfying δ. We then apply an exit strategy that suppresses the release of the real positions following v_{i-1}, simulating u stopping at v_{i-1}.

Note that the approach described in this section is used only during the learning phase. In fact, in the transition function associated with u, sensitive movements will be associated with a probability equal to 0, and then managed as unusual paths.

Figure 5 illustrates an example of how inferences on sensitive movements are counteracted. Figure 5(a) shows the real path of Bob going from work to Milan downtown. Given $\lambda(\mathtt{w},\mathtt{d})$=1, the middleware computes E_{safe}. In our example, E_{safe}=$\emptyset$, since no positions in V are closer than δ=200 meters to Milan downtown. The middleware then creates a cover story in which Bob is at work. Figure 5(b) illustrates the cover story released to the LBS provider.

Inference on unusual paths. When a user u follows an unusual path $\langle v_1,\ldots,v_{i-1},v_i\rangle$, a risk of inference arises and the path should be manipulated before its release. Our solution builds a cover story in which the path followed by u is obfuscated by the middleware to release a different path $\langle v_1,\ldots,v_{i-1},v_j\rangle$, which can be considered as usual.

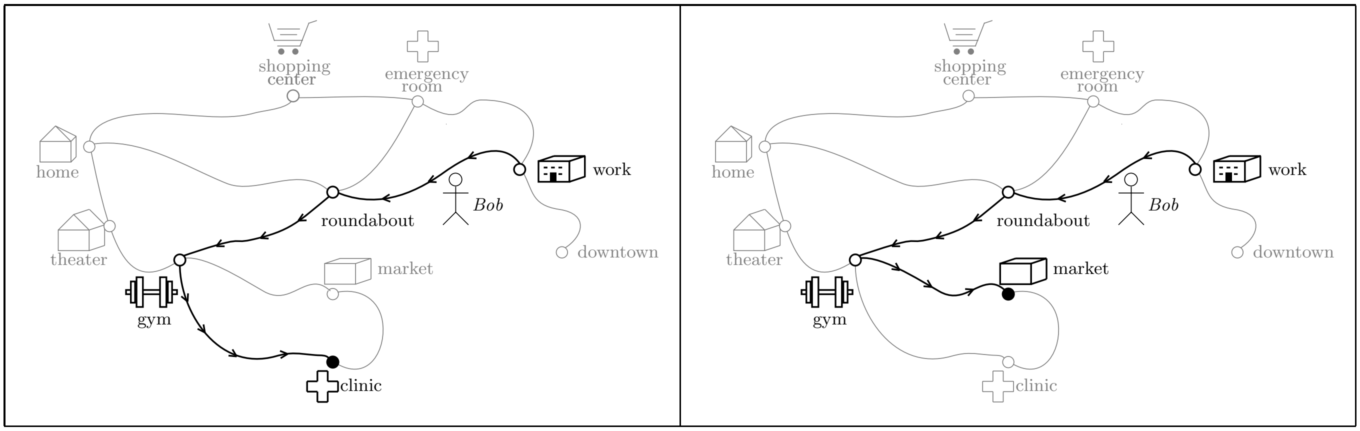
The first step in this scenario requires to identify an unusual path, based on the user behavior captured by the transition



(a) Real path

(b) Released path

Fig. 5: Counteracting inference on sensitive movements



(a) Real path

(b) Released path

Fig. 6: Counteracting inference on unusual paths

function T. In this paper, we consider unusual those paths whose probability (Definition 5.2) deviates from an average probability. In particular, given a position v_{i-1} in the path at a given point in time t, we evaluate the usualness of the path taken by the user toward v_i with respect to the possible edges (v_{i-1}, v_j) that u can choose at time $t{+}1$. An unusual path can then be formally defined as follows.

Definition 6.3 (Unusual path): Let u be a user who has released a path $\langle v_1,\ldots,v_{i-1}\rangle$ and is moving from v_{i-1} to v_i. The path $\langle v_1,\ldots,v_{i-1},v_i\rangle$ is *unusual* if $P(\langle v_1,\ldots,v_{i-1},v_i\rangle){<}\alpha{\cdot}\frac{\sum_{v_j\in V_j} P(\langle v_1,\ldots,v_{i-1},v_j\rangle)}{|V_j|}$, with $V_j{=}\{v_j|\exists(v_{i-1},v_j)$ and $d(v_i,v_j){\leq}\delta\}$ and $\alpha{\in}[0,1]$.

We note that α is a weight applied to the average probability of all paths $\langle v_1,\ldots,v_{i-1},v_j\rangle$ that should be considered in the discovery of unusual path. The higher the value for α, the more restrictive the condition that a path must satisfy to be considered usual. If the probability $P(\langle v_1,\ldots,v_{i-1},v_i\rangle)$ exceeds the average value weighted by α, then the path leading to v_i is usual.

Similarly to Definition 6.2, in case the path is unusual, the privacy middleware releases a cover story using a greedy approach. It then selects the usual path $\langle v_1,\ldots,v_{i-1},v_j\rangle$ with higher probability such that $d(v_i,v_j){\leq}\delta$.

Figure 6 illustrates an example of how inferences on unusual paths are counteracted. Figure 6(a) shows the real path of Bob going from work to the clinic for a different medical advice. According to the transition function T in Figure 3(b), the path $\langle\mathtt{w},\mathtt{r},\mathtt{g}\rangle$ can be considered usual and, therefore, the middleware does not build any cover story for it. However, leaving the gym toward the clinic (i.e., $(\mathtt{g},\mathtt{c})$) makes the path $\langle\mathtt{w},\mathtt{r},\mathtt{g},\mathtt{c}\rangle$ unusual. Therefore, according to T and δ, the middleware builds the cover story $\langle\mathtt{w},\mathtt{r},\mathtt{g},\mathtt{m}\rangle$. Figure 6(b) illustrates the cover story released to the LBS provider.

VII. Related Work

The problem of protecting location privacy has been under the attention of many researchers in recent years, resulting in a large number of privacy-enhancing solutions [1], [2].

First approaches to the protection of location privacy have considered a scenario in which a single position is released to the LBS. In this context two main classes of techniques have been defined: anonymity-based and obfuscation-based. Anonymity-based solutions (e.g., [2], [5], [6], [7], [8], [9],

[10]) aim at protecting the association between a user and her sensitive information by avoiding the possibility to re-identify the user observing her request(s). Such techniques are typically based on the concept of k-anonymity, originally devised for databases and data release scenarios [11]. The general idea is that of degrading the precision of the queries posed by anonymized users in such a way that at least k different users are indistinguishable from their location. Obfuscation-based techniques (e.g., [3], [12], [13]) aim at protecting location privacy by degrading the accuracy of users' location. The main goal of these techniques is to perturb the location information of the users still maintaining a binding with their identity.

Recently, a line of work has addressed the problem of protecting path, or trajectory, privacy [1]. The solution proposed in [14] is based on a spatial cloaking technique, introducing the concept of *dynamic* grouping of users issuing queries. This technique ensures that a cloaked spatial region is shared by at least k users and, to protect user trajectories, that all k users appear as belonging to the same region as time passes. In [15], the authors put forward the idea of evaluating the distortion associated with location information by means of the perimeter of a cloaked region. In [16], the protection level characterizing a circular cloaked spatial region is measured in terms of its entropy, and a polynomial-time algorithm is proposed to determine the minimum circle that surrounds a user and other $k-1$ users. A different solution is proposed in [17], where first-order Markov chains are used to predict synthetic trajectories from historical data, and intersecting paths are submitted at the same time to the service provider to guarantee the creation of a mix-zone. Releasing simulated locations is at the basis of the technique proposed in [18]. This technique adopts probabilistic models of driving behaviors, applied for creating realistic driving trips, and GPS noise to decrease the precision of the starting point of a trip. A method for preserving privacy of GPS traces guaranteeing an appropriate protection level to users moving in low-density areas is proposed in [19]. All the above solutions provide mechanisms to protect the location privacy of the users at different levels, but do not address the problem of protecting users' from inferences done on the released paths. Solutions for path privacy protection in fact mostly focus on guaranteeing the anonymity of the users. By providing a solution that reduces inferences on released paths, we make a new step toward the definition of a privacy-preserving mobile environment.

VIII. Conclusions and Future Work

Mobile communications and location-based services are playing a crucial role in today IT systems. However, the unrestricted management of related location information is putting at risk the privacy of the users and may result in a society where mobile technologies – whose primary goal is to enable the development of innovative and valuable services – are used to track and keep individuals under control. The approach presented in this paper aims at providing a solution that limits the amount of inference that can be drawn by an attacker observing users' movements. Our approach can be easily integrated within existing devices, presents a modeling of users' movements based on Markov chains, and provides a simple solution for expressing and enforcing users' preferences. Possible future work includes an enhanced solution for modeling users' behavior, a smarter selection of cover stories, and an approach to evaluate the quality of the obfuscated paths.

Acknowledgments

This work was supported by the Italian Ministry of Research within the PRIN 2008 project "PEPPER" (2008SY2PH4).

References

[1] C.-Y. Chow and M. F. Mokbel, "Trajectory privacy in location-based services and data publication," *SIGKDD Explorations Newsletter*, vol. 13, no. 1, pp. 19–29, August 2011.

[2] C. Bettini, S. Jajodia, P. Samarati, and X. S. Wang, Eds., *Privacy in Location-Based Applications: Introduction, Research Issues and Applications*. LNCS 5599, Springer, 2009, vol. 5599.

[3] C. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati, "An obfuscation-based approach for protecting location privacy," *IEEE TDSC*, vol. 8, no. 1, pp. 13–27, January-February 2011.

[4] B. Everitt, *The Cambridge dictionary of statistics*, 2nd ed. Cambridge University Press, Cambridge, U.K., 2002.

[5] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. of MobiSys 2003*, San Francisco, CA, USA, May 2003.

[6] M. Mokbel, C.-Y. Chow, and W. Aref, "The new Casper: query processing for location services without compromising privacy," in *Proc. of VLDB 2006*, Seoul, Korea, September 2006.

[7] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE TMC*, vol. 7, no. 1, pp. 1–18, January 2008.

[8] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries," *IEEE TKDE*, vol. 19, no. 12, pp. 1719–1733, December 2007.

[9] K. Mouratidis and M. Yiu, "Anonymous query processing in road networks," *IEEE TKDE*, vol. 22, no. 1, pp. 2–15, January 2010.

[10] C. Ardagna, S. Jajodia, P. Samarati, and A. Stavrou, "Providing mobile users' anonymity in hybrid networks," in *Proc. of ESORICS 2010*, Athens, Greece, September 2010.

[11] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, and P. Samarati, "k-Anonymity," in *Secure Data Management in Decentralized Systems*, T. Yu and S. Jajodia, Eds. Springer, 2007.

[12] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," in *Proc. of PERVASIVE 2005*, Munich, Germany, May 2005.

[13] C. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati, "Location privacy protection through obfuscation-based techniques," in *Proc. of DBSec 2007*, Redondo Beach, CA, USA, July 2007.

[14] C.-Y. Chow and M. Mokbel, "Enabling private continuous queries for revealed user locations," in *Proc. of SSTD 2007*, Boston, MA, USA, July 2007.

[15] X. Pan, X. Meng, and J. Xu, "Distortion-based anonymity for continuous queries in location-based mobile services," in *Proc. of ACM GIS 2009*, Seattle, WA, USA, November 2009.

[16] T. Xu and Y. Cai, "Location anonymity in continuous location-based services," in *Proc. of ACM GIS 2007*, Seattle, WA, USA, Nov. 2007.

[17] J. Meyerowitz and R. Roy Choudhury, "Hiding stars with fireworks: location privacy through camouflage," in *Proc. of MobiCom 2009*, Beijing, China, September 2009.

[18] J. Krumm, "Realistic driving trips for location privacy," in *Proc. of Pervasive 2009*, Nara, Japan, May 2009.

[19] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, "Preserving privacy in GPS traces via uncertainty-aware path cloaking," in *Proc. of CCS 2007*, Alexandria, VA, USA, October-November 2007.

[20] S. Martinez-Bea and V. Torra, "An evaluation framework for location privacy," in *Proc. of CCIA 2011*, Lleida, Catalunia, Spain, October 2011.

[21] C. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati, "Supporting location-based conditions in access control policies," in *Proc. of ASIACCS 2006*, Taipei, Taiwan, March 2006.