Limits of Location Privacy under Anonymization and Obfuscation

Nazanin Takbiri
Electrical and
Computer Engineering
UMass-Amherst
ntakbiri@umass.edu

Amir Houmansadr
Information and
Computer Sciences
UMass-Amherst
amir@cs.umass.edu

Dennis L. Goeckel
Electrical and
Computer Engineering
UMass-Amherst
goeckel@ecs.umass.edu

Hossein Pishro-Nik
Electrical and
Computer Engineering
UMass-Amherst
pishro@ecs.umass.edu

Abstract—**The prevalence of mobile devices and location-based services (LBS) has generated great concerns regarding the LBS users' privacy, which can be compromised by statistical analysis of their movement patterns. A number of algorithms have been proposed to protect the privacy of users in such systems, but the fundamental underpinnings of such remain unexplored. Recently, the concept of perfect location privacy was introduced and its achievability was studied for anonymization-based LBS systems, where user identifiers are permuted at regular intervals to prevent identification based on statistical analysis of long time sequences. In this paper, we significantly extend that investigation by incorporating the other major tool commonly employed to obtain location privacy: obfuscation, where user locations are purposely obscured to protect their privacy. Since anonymization and obfuscation reduce user utility in LBS systems, we investigate how location privacy varies with the degree to which each of these two methods is employed. We provide: (1) achievability results for the case where the location of each user is governed by an i.i.d. process; (2) converse results for the i.i.d. case as well as the more general Markov Chain model. We show that, as the number of users in the network grows, the obfuscation-anonymization plane can be divided into two regions: in the first region, all users have perfect location privacy; and, in the second region, no user has location privacy.**

Index Terms—**Location Based Service (LBS), Location Privacy Protecting Mechanism (LPPM), Information Theoretic Privacy, Anonymization, Obfuscation, Markov chain.**

I. Introduction

Mobile devices, ranging from smart phones to connected automobiles, offer a wide spectrum of location-based services (LBS), such as ride sharing, navigation, dining recommendations, and accident warnings. However, these important services can cause significant privacy threats to their users, as even anonymized time series of locations can be statistically matched to prior user behavior to allow for user identification and tracking. Therefore, privacy in LBS applications is an important and difficult challenge.

There are two major types of location privacy protection mechanisms (LPPMs): identity perturbation (anonymization) techniques, and location perturbation (obfuscation) techniques. In anonymization techniques, privacy is obtained by concealing the mapping between users and location observations, and the mapping is changed periodically to attempt to thwart statistical analysis, which benefits from long time series when matching anonymized traces to prior user behavior. In obfuscation techniques, privacy is obtained by returning purposefully inaccurate (i.e., noisy) location information to the LBS applications.

Anonymization and obfuscation improve user privacy at the cost of user utility. In anonymization, we need to change these pseudonym mappings frequently to achieve high privacy by reducing the length of time series exploited by statistical analysis. However, this frequent change could decrease usability and functionality by concealing the temporal relation between a user's locations, which may be critical in the utility of some LBS systems, e.g., a dining recommendation system that makes suggestions based on the dining places visited by a user in the past. For obfuscation-based mechanisms, the added noise to the reported values of user locations will degrade the utility of LBS applications that are sensitive to the absolute values of location information, e.g., a ride-sharing LBS system. Thus, choosing the right level of privacy-protection mechanism is an important question, and understanding what levels of anonymization and obfuscation can provide theoretical guarantees of privacy is of interest.

Despite extensive previous studies on location privacy and LPPM mechanisms, the theoretical foundations of location privacy have not yet been established. In [1]–[4], an approach was introduced to understand the fundamental limits of location privacy when only anonymization is used. There, users are characterized by the statistics of their locations, and the adversary then tries to match traces to those statistics to de-anonymize users. Per above, anonymization thwarts such statistical analysis by reducing the time series available for such a matching, and thus [1]–[3] consider the rate at which pseudonyms must be changed so as to preserve perfect location privacy. In this paper, the perfect privacy set-up of [3] is again employed; however, in addition to anonymization, location perturbation (obfuscation) is also considered; thus, the adversary attempts to infer information about the actual locations by observing the obfuscated *and* anonymized version of the location data. We study both achievability and converse results.

Due to space limitations, the proofs are provided in the long version of the paper [5].

This work was supported by National Science Foundation under grants CCF– 0844725 and CCF–1421957.

II. Related Work

Prior LPPM mechanisms can be categorized into two main groups: identity perturbation LPPMs (anonymization) [6], [7] and location perturbation LPPMs (obfuscation) [8], [9]. Some prior works combine techniques to achieve stronger location privacy. For instance, Freudiger et al. combine techniques from cryptography with mix-zones to improve location privacy [10]. Differential privacy [11] has also been applied to the problem of location privacy [12], [13].

Several studies aim at quantifying location privacy protection of specific LPPMs [14]–[16]. To defeat a localization attack and achieve privacy at the same time, [17] proposed a method which finds the optimal LPPM for an LBS given service quality constraints.

However, the literature is missing a theoretical framework that would allow us to provide provable location privacy guarantees, obtain fundamental trade-offs between location privacy and utility, and provide tools to optimally achieve location privacy. Here, we continue to employ our metric based on mutual information that was introduced in [1]–[3].

Mutual information has been used by others as a privacy metric in various settings [18]–[22]. However, much of this work is motivated by differential privacy, and thus prior work is not directly applicable to the location privacy problem: large sets of time-series data belonging to different users with different movement dynamics that has gone through an LPPM.

Unnikrishnan provides a comprehensive analysis of asymptotically optimal matching of time series to source distributions [23]. However, he does not study any privacy metrics as considered in this paper. In fact, the most difficult technical challenges we faced were in showing that the mutual information converges to zero so we can conclude there is no privacy leakage. There are also significant differences in the setting, as: (1) [23] doesn't study obfuscation; (2) [23] does not consider non-i.i.d. cases; (3) fitting our application, we assume the existence of a general (but possibly unknown) prior distribution for the sources (i.e., a Bayesian setting); and (4) we study the asymptotic limits in terms of both the number of users and the number of observations.

III. Framework

We assume a system with n users. $X_u(k)$ denotes the location of user u at time k, which would we like to protect from an interested adversary $\mathcal{A}$. To study guarantees of privacy, we desire to protect against a powerful adversary, and thus we assume the adversary $\mathcal{A}$ has a complete statistical model of the users' movements. Figure 1 shows the LPPM configuration considered here: $Z_u(k)$ shows the (reported) location of user u at time k after applying obfuscation to $X_u(k)$, and $Y_u(k)$ shows the (reported) location after applying anonymization to the data. As in [1]–[3], we assume that the anonymization permutes user pseudonyms every $m(n)$ observations. Hence, for attempting to determine the locations $X_u(k), k = 1, 2, \dots, m(n)$, the adversary employs the observations of all users from times $k = 1, 2, \cdots, m(n)$.

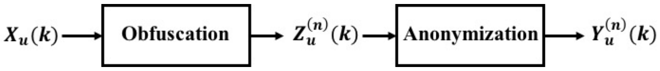

Fig. 1: The adversary $\mathcal{A}$ attempts to estimate $X_u(k), k = 1, 2, \dots, m(n)$ from $Y_u^{(n)}(k), k = 1, 2, \dots, m(n)$.

Let $\mathbf{X}_u^{(n)}$ be the vector which contains $m(n)$ locations of user u, and $\mathbf{X}^{(n)}$ is the $m(n) \times n$ matrix which contains $\mathbf{X}_u^{(n)}$ for all users,

$$\mathbf{X}_u^{(n)} = \begin{bmatrix} X_u(1) \\ X_u(2) \\ \vdots \\ X_u(m) \end{bmatrix}, \quad \mathbf{X}^{(n)} = \left[\mathbf{X}_1^{(n)}, \mathbf{X}_2^{(n)}, \cdots, \mathbf{X}_n^{(n)}\right].$$

Location Data Model: Assume the users move over $r \geq 2$ possible locations $(0, 1, \cdots, r-1)$. At any time, $X_u(k)$ is equal to a value in $\{0, 1, \cdots, r-1\}$ according to that user's probability distribution, which satisfies some mild regularity conditions given below. Per above, the users' distributions are known to the adversary $\mathcal{A}$, and he/she employs them to distinguish different users based on traces of user activity of length $m(n)$. In the first part of the paper, we assume an i.i.d. model; each user's locations at different times are drawn independently from their distribution, as would be true in situations where the location data is sampled at a low rate. Any dependency can only favor the adversary, so these results also provide lower bounds on the achievable privacy in those settings, which would be desirable if the dependency is poorly understood. Later, we consider the case where the movements are modeled by Markov chains (Section V-C).

Obfuscation Model: We assume that each user has only limited knowledge of the characteristics of the overall population and thus we employ a simple distributed method in which the location of each user is reported with error with a certain probability, where that probability itself is generated randomly for each user. More precisely, let $\mathbf{Z}_u^{(n)}$ be the vector which contains the obfuscated versions of user u's locations, and $\mathbf{Z}^{(n)}$ is the collection of $\mathbf{Z}_u^{(n)}$ for all users,

$$\mathbf{Z}_u^{(n)} = \begin{bmatrix} Z_u^{(n)}(1) \\ Z_u^{(n)}(2) \\ \vdots \\ Z_u^{(n)}(m) \end{bmatrix}, \quad \mathbf{Z}^{(n)} = \left[\mathbf{Z}_1^{(n)}, \mathbf{Z}_2^{(n)}, \cdots, \mathbf{Z}_n^{(n)}\right].$$

For user u, we generate the random variable $R_u^{(n)}$ that is uniformly distributed between 0 and a_n[1]. The value of $R_u^{(n)}$ is the probability that any location of the user is changed to a different location by obfuscation, and a_n is termed the "noise level" of the system. Hence, the effect of the obfuscation is to alter the probability distribution function of each user across the r locations in a way that is unknown to the adversary,

[1] The uniform distribution assumption for $R_u^{(n)}$ is not necessary and the results can be extended to a general set of distributions when we employ $a_n = E[R_u^{(n)}]$. However, the assumption of a uniform random variable simplifies the presentation of the results significantly.

since the obfuscation is independent of all past activity of the user.

Anonymization Model: Anonymization is modeled by a random permutation $\Pi^{(n)}$ on the set of n users. The user u is assigned the pseudonym $\Pi^{(n)}(u)$. For simplicity, we will employ $\Pi(u)$ instead of $\Pi^{(n)}(u)$ where the meaning is not ambiguous. $\mathbf{Y}^{(n)}$ is the anonymized version of $\mathbf{Z}_u^{(n)}$; thus

$$\begin{aligned}\mathbf{Y}^{(n)} &= \text{Perm}\left(\mathbf{Z}_1^{(n)}, \mathbf{Z}_2^{(n)}, \cdots, \mathbf{Z}_n^{(n)}; \Pi^n\right) \\ &= \left[\mathbf{Z}_{\Pi^{-1}(1)}^{(n)}, \mathbf{Z}_{\Pi^{-1}(2)}^{(n)}, \cdots, \mathbf{Z}_{\Pi^{-1}(n)}^{(n)}\right] \\ &= \left[\mathbf{Y}_1^{(n)}, \mathbf{Y}_2^{(n)}, \cdots, \mathbf{Y}_n^{(n)}\right],\end{aligned}$$

where Perm(. , Π) shows the permutation operation using the permutation function Π. As a result, $\mathbf{Y}_u^{(n)} = \mathbf{Z}_{\Pi^{-1}(u)}^{(n)}$ and $\mathbf{Y}_{\Pi(u)}^{(n)} = \mathbf{Z}_u^{(n)}$.

Adversary Model: The adversary is assumed to have complete statistical knowledge of the users' movements. The adversary also knows the value of a_n, as it is a design parameter. However, the adversary does not know the realization of the random permutation $\Pi^{(n)}$ or the realizations of random variables $R_u^{(n)}$, as these are independent of the past behavior of the users.

Perfect location privacy is defined as follows [3]:

Definition 1. User u has *perfect location privacy* at time k, if and only if

$$\lim_{n\to\infty} I\left(X_u(k); \mathbf{Y}^{(n)}\right) = 0, \tag{1}$$

In this paper, we also consider the situation in which there is no location privacy:

Definition 2. User u has *no location privacy* at time k, if there exists an algorithm for the adversary to estimate $X_u(k)$ such that, as $n \to \infty$,

$$P_e(u) \triangleq P\left(\widetilde{X_u(k)} \neq X_u(k)\right) \to 0,$$

where $\widetilde{X_u(k)}$ is the estimated location of user u at time k.

IV. Perfect Privacy: Achievability

A. Two-State Model

We first consider the two-state ($r = 2$) model, which captures the salient aspects of the problem, in particular the proof techniques that will be employed throughout. With two possible locations 0 and 1, a user's location is characterized by a Bernoulli random variable with a single parameter p_u, which is the probability of user u being at location 1. Thus, $X_u(k) \sim Bernoulli\left(p_u\right)$. We assume that p_u, $u = 1, 2, \ldots, n$ are drawn independently from a continuous density function, $f_P(p_u)$, on the $(0, 1)$ interval. Specifically, we assume there are $\delta_1, \delta_2 > 0$ such that[2]:

$$\begin{cases}\delta_1 < f_P(p_u) < \delta_2 & p_u \in (0,1) \\ f_P(p_u) = 0 & p_u \notin (0,1)\end{cases}$$

Since $X_u(k)$ are i.i.d. and have a Bernoulli distribution, $Z_u(k)$'s and $Y_u(k)$'s are also i.i.d. with a Bernoulli distribution.

The adversary, armed with knowledge of $p_u, u = 1, 2, \ldots, n$, and the obfuscated and anonymized observations $Y_u(k)$ for $u = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, m(n)$, wants to determine the locations $X_u(k), k = 1, 2, \ldots, m(n)$. The following theorem states that if a_n is significantly larger than $\frac{1}{n}$ in this two-state model, then all users have perfect location privacy independent of the value of $m(n)$.

Theorem 1. For the above two-state model, if

- $m = m(n)$ is arbitrary;
- $R_u^{(n)} \sim Uniform[0, a_n]$, where $a_n \triangleq c' n^{-(1-\beta)}$ for $c' > 0$ and $\beta > 0$;

then user 1 has perfect location privacy. That is,

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} I\left(X_1(k); \mathbf{Y}^{(n)}\right) = 0.$$

Proof: Recall that the proofs of all theorems are in [5]. ■

By symmetry, the theorem readily applies to all users $u = 1, 2, \ldots, n$.

Although $m(n)$ is arbitrary and thus can be arbitrarily large, suggesting anonymization becoming unnecessary in the limit, it is important to note that some (small) degree of anonymization is indeed required. In particular, asymptotically large $m(n)$ does imply that the system does not need to change the pseudonyms during system operation. However, the system does need to assign pseudonyms once to the users, as it is the obfuscation preventing the adversary from determining which trace is associated with user u that prevents him from tracking user u in the limit of large n. This is readily observed by noting that, without anonymization, the noise level of the obfuscation specified in Theorem 1 goes to zero for large n; hence, if the adversary knows which trace is associated with user u, that user has no location privacy.

B. Extension to r-States

Now, assume an r-location model with locations $(0, 1, \cdots, r-1)$, where $p_u(i)$ shows the probability of user u being at location i. We define the vector $\mathbf{p}_u$

$$\mathbf{p}_u = \left[p_u(1), p_u(2), \cdots, p_u(r-1)\right]^T.$$

We assume $p_u(i)$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, on the $(0, 1)^{r-1}$ hypercube (note that the $p_u(i)$'s sum to one, so one of them can be considered as the dependent value and the dimension is $r-1$). In particular, define the range of the distribution as

$$\begin{aligned}R_\mathbf{P} = \{(x_1, \cdots, x_{r-1}) \in (0,1)^{r-1} : \\ x_i > 0, x_1 + x_2 + \cdots + x_{r-1} < 1\}.\end{aligned}$$

Then, we assume there are $\delta_1, \delta_2 > 0$ such that:

$$\begin{cases}\delta_1 < f_\mathbf{P}(\mathbf{p}_u) < \delta_2, & \mathbf{p}_u \in R_\mathbf{P} \\ f_\mathbf{P}(\mathbf{p}_u) = 0, & \mathbf{p}_u \notin R_\mathbf{P}\end{cases}$$

The obfuscation is similar to the two-state case. Specifically, for $j \in \{0, 1, \cdots, r-1\}$ we can write

[2] The condition $\delta_1 < f_P(p_u) < \delta_2$ is not actually necessary for the results and can be relaxed; however, we keep it here to avoid unnecessary technicalities.

$$P(Z_u^{(n)}(k) = j|X_u(k) = i) = \begin{cases} 1 - R_u^{(n)}, & \text{for } j = i \\ \frac{R_u^{(n)}}{r-1}, & \text{for } j \neq i \end{cases}$$

Theorem 2. For the above r-state model, if:

- $m = m(n)$ is arbitrary;
- $R_u^{(n)} \sim Uniform[0, a_n]$, where $a_n \triangleq c'n^{-\left(\frac{1}{r-1}-\beta\right)}$ for $c' > 0$ and $\beta > 0$;

then user 1 has perfect location privacy. That is,

$$\forall k \in \mathbb{N}, \quad \lim_{n\to\infty} I\left(X_1(k); \mathbf{Y}^{(n)}\right) = 0.$$

V. Converse Results: No Privacy

In this section, we prove that if the number of observations by the adversary is larger than its critical value and the value of the noise level is less than its critical value, then the adversary can find an algorithm to successfully estimate the location of users with arbitrarily small error probability. Combined with the results of the previous section, this implies that asymptotically (as $n \to \infty$), location privacy can be achieved *if and only if* at least one of the two techniques (obfuscation or anonymization) are used on the proper side of their thresholds. However, it is important to recall the discussion following Theorem 1. In particular, looking at the results of [3], we notice that anonymization alone can provide perfect privacy if $m(n)$ is below its threshold (even when no obfuscation is employed). On the other hand, as described in the discussion after Theorem 1, the threshold for obfuscation is only valid when obfuscation is used in conjunction with anonymization (with arbitrary $m(n)$) as shown in Figure 1.

A. Two-State Model

Consider first the i.i.d. two-state model. Recall that we can consider the location of users at any time as a Bernoulli random variable with parameter p_u. As before, we assume that p_u's are drawn independently from some continuous density function, $f_P(p_u)$, on the $(0, 1)$ interval. Specifically, there are $\delta_1, \delta_2 > 0$ such that

$$\begin{cases} \delta_1 < f_P(p_u) < \delta_2, & p_u \in (0,1) \\ f_P(p_u) = 0, & p_u \notin (0,1) \end{cases}$$

Theorem 3. For the two-state i.i.d. model, if:

- $m = cn^{2+\alpha}$ for $c > 0$ and $\alpha > 0$;
- $R_u^{(n)} \sim Uniform[0, a_n]$, where $a_n \triangleq c'n^{-(1+\beta)}$ for $c' > 0$ and $\beta > \frac{\alpha}{4}$;

then user 1 has no location privacy as n goes to infinity. In other words, there exists an algorithm for the adversary to estimate $X_1(k)$ such that

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0 \quad \text{as } n \to \infty.$$

Note that due to the symmetry of the problem, the theorem applies to all users.

The basic idea is that the adversary first inverts the anonymization mapping $\Pi(n)$ to obtain $Z_1(k)$, and then estimates the value of $X_1(k)$ from that. To invert the anonymization, the adversary calculates the empirical averages for the observed presence of users at location 1 and then assigns the string with the empirical average closest to p_1 to user 1.

B. Extension to r-di.i.dStates

Now, assume users can go to r locations $(0, 1, \cdots, r-1)$, with $p_u(i)$ the probability of user u being at location i. The vector $\mathbf{p}_u$ is defined as in Section IV-B. We also consider $p_u(i)$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, on the $(0, 1)^{r-1}$ hypercube. We also defined $f_P(\mathbf{p}_u)$ and the range of distribution in Section IV-B.

Theorem 4. For the above r-state mode, if:

- $m = cn^{\frac{2}{r-1}+\alpha}$ for $c > 0$ and $\alpha > 0$;
- $R_u^{(n)} \sim Uniform[0, a_n]$, where $a_n \triangleq c'n^{-\left(\frac{1}{r-1}+\beta\right)}$ for $c' > 0$ and $\beta > \frac{\alpha}{4}$;

then user 1 has no location privacy as n goes to infinity. In other words,

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0.$$

C. Markov Chain Model

To this point, we have assumed there are r locations and users' movements are i.i.d. Here, we model users' movements by Markov chains to capture the dependency of the users' movement across time. Again, we assume there are r possible locations, which in this case corresponds to the number of states in the Markov chain. Let E be the set of edges. More specifically, $(i, j) \in E$ if there exists an edge from i to j with probability $p(i, j) > 0$. In this case, the users are distinguished by their transition probabilities $p_u(i, j)$ (where subscript u refers to user u). The adversary $\mathcal{A}$ again knows the transition probabilities of all users, and the model for obfuscation and anonymization is exactly the same as before.

We show that the adversary will be able to estimate the locations of the users with low error probability if $m(n)$ and a_n are in the appropriate range. The key idea is that the adversary can focus on a subset of transition probabilities that are sufficient for recovering the entire transition probability matrix. By estimating those transition probabilities from the observed data and matching them with the known transition probabilities of the users, the adversary will be able to first de-anonymize the data, and then estimate the locations of users.

Theorem 5. For an irreducible, aperiodic Markov chain with r states and $|E|$ edges as defined above, if:

- $m = cn^{\frac{2}{|E|-r}+\alpha}$ for $c > 0$ and $\alpha > 0$;
- $R_u^{(n)} \sim Uniform[0, a_n]$, where $a_n \triangleq c'n^{-\left(\frac{1}{|E|-r}+\beta\right)}$ for $c' > 0$ and $\beta > \frac{\alpha}{4}$;

then the adversary can successfully identify the location of user 1 as n goes to infinity. In other words,

$$P_e(1) \triangleq P\left(\widetilde{X_1(k)} \neq X_1(k)\right) \to 0.$$

VI. Conclusion

In this paper, we have considered both obfuscation and anonymization techniques to achieve location privacy from an information-theoretic perspective. In particular, we have employed the mutual information metric from [3] to explore the minimum amounts of obfuscation and anonymization such that there is no information leakage to an interested adversary who possesses full statistical knowledge of user mobility patterns. We have characterized the limits of location privacy in the entire $m(n) - a_n$ plane for the i.i.d. case, as shown in Figure 2. The privacy level of the users depends on both $m(n)$ (number of observations per user by the adversary for a fixed anonymization mapping) and a_n (noise level of the obfuscation). That is, larger $m(n)$ and smaller a_n indicate weaker location privacy. In this paper, we obtained the exact values of the thresholds for $m(n)$ and a_n. We showed that if $m(n)$ is fewer than $O\left(n^{\frac{2}{r-1}}\right)$, or a_n is bigger than $\Omega\left(n^{-\frac{1}{r-1}}\right)$, users have perfect location privacy. On the other hand, if none of the above conditions are satisfied, users have no location privacy. For the case where the users' movements are modeled by Markov chains, we obtained a no-privacy region in the $m(n) - a_n$ plane.

It is worth noting that these are somewhat coarse-grained thresholds in the sense that we have proven the results up to the exponents of n. In other words, we did not investigate the case that we are exactly at the thresholds or maybe a factor $\log n$ above or below the thresholds. Such a fine-grained investigation left for the future work.

Future research in this area needs to characterize the exact privacy/no-privacy regions under Markov models for user movements. It is also important to consider different ways to obfuscate users' movements, and study the utility-privacy trade-offs across these different obfuscation techniques.

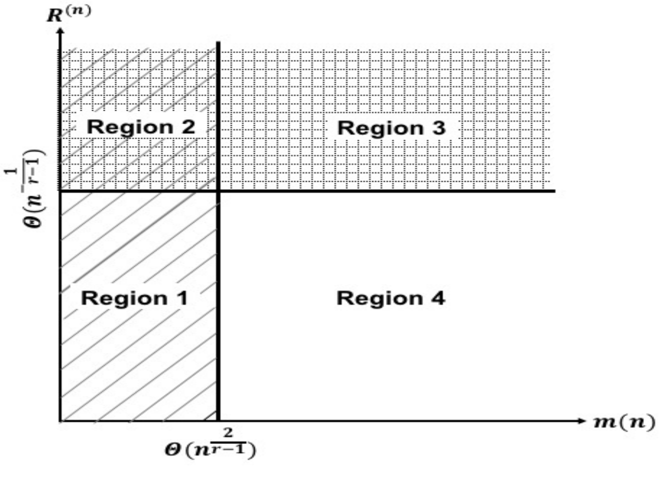

Fig. 2: Limits of location privacy in the $m(n) - a_n$ plane for the i.i.d. case: in regions 1, 2, and 3, users have perfect location privacy, and in region 4, users have no location privacy.

References

[1] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Defining perfect location privacy using anonymization," in *2016 Annual Conference on Information Science and Systems (CISS)*. IEEE, 2016, pp. 204–209.

[2] Z. Montazeri, A. Houmansadr, and H.Pishro-Nik, "Achieving perfect location privacy in markov models using anonymization," in *2016 International Symposium on Information Theory and its Applications (ISITA2016)*, Monterey, USA, oct 2016.

[3] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Achieving Perfect Location Privacy in Wireless Devices Using Anonymization," *under revision in IEEE Transactions on Information Forensics and Security*, 2016.

[4] N. Takbiri, A. Houmansadr, D. Goeckel, and H. Pishro-Nik, "Fundamental limits of location privacy using anonymization," in *Annual Conference on Information Science and Systems (CISS)*. IEEE, 2017.

[5] N. Takbiri, A. Houmansadr, D. L. Goeckel, and H. Pishro-Nik, "Matching Anonymized and Obfuscated Time Series to Users' Profile," *to be submitted in IIEEE Transactions on Information Theory*, 2017.

[6] G. P. Corser, H. Fu, and A. Banihani, "Evaluating location privacy in vehicular communications and applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 9, pp. 2658–2667, 2016.

[7] F. Naini, J. Unnikrishnan, P. Thiran, and M. Vetterli, "Where you are is who you are: User identification by matching statistics," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 358–372, 2016.

[8] R. Shokri, G. Theodorakopoulos, P. Papadimitratos, E. Kazemi, and J.-P. Hubaux, "Hiding in the mobile crowd: Locationprivacy through collaboration," *IEEE transactions on dependable and secure computing*, vol. 11, no. 3, pp. 266–279, 2014.

[9] H. Kido, Y. Yanagisawa, and T. Satoh, "Protection of location privacy using dummies for location-based services," in *Data Engineering Workshops, 2005. 21st International Conference on*. IEEE, 2005, pp. 1248–1248.

[10] J. Freudiger, M. Raya, M. Félegyházi, P. Papadimitratos, and J.-P. Hubaux, "Mix-zones for location privacy in vehicular networks," in *CM Workshop on Wireless Networking for Intelligent Transportation Systems (WiN-ITS)*, 2007.

[11] J. Lee and C. Clifton, "Differential identifiability," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1041–1049.

[12] R. Dewri, "Local differential perturbations: Location privacy under approximate knowledge attackers," *Mobile Computing, IEEE Transactions on*, vol. 12, no. 12, pp. 2360–2372, 2013.

[13] 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*. ACM, 2013, pp. 901–914.

[14] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Security and Privacy (SP), 2011 IEEE Symposium on*. IEEE, 2011, pp. 247–262.

[15] R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, "Quantifying location privacy: the case of sporadic location exposure," in *Privacy Enhancing Technologies*. Springer, 2011, pp. 57–76.

[16] Z. Ma, F. Kargl, and M. Weber, "A location privacy metric for v2x communication systems," in *Sarnoff Symposium, 2009. SARNOFF'09. IEEE*. IEEE, 2009, pp. 1–6.

[17] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: optimal strategy against localization attacks," in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012, pp. 617–627.

[18] S. Salamatian, A. Zhang, F. du Pin Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft, "How to hide the elephant-or the donkey-in the room: Practical privacy against statistical inference for large data." in *GlobalSIP*, 2013, pp. 269–272.

[19] I. Csiszár, "Almost independence and secrecy capacity," *Problemy Peredachi Informatsii*, vol. 32, no. 1, pp. 48–57, 1996.

[20] F. P. Calmon, A. Makhdoumi, and M. Médard, "Fundamental limits of perfect privacy," in *Information Theory (ISIT), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 1796–1800.

[21] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 6, pp. 838–852, 2013.

[22] H. Yamamoto, "A source coding problem for sources with additional outputs to keep secret from the receiver or wiretappers (corresp.)," *IEEE Transactions on Information Theory*, vol. 29, no. 6, pp. 918–923, 1983.

[23] J. Unnikrishnan, "Asymptotically optimal matching of multiple sequences to source distributions and training sequences," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 452–468, 2015.