Privacy against Statistical Matching: Inter-User Correlation

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*—Modern applications significantly enhance user experience by adapting to each user's individual condition and/or preferences. While this adaptation can greatly improve utility or be essential for the application to work (e.g., for ride-sharing applications), the exposure of user data to the application presents a significant privacy threat to the users, even when the traces are anonymized, since the statistical matching of an anonymized trace to prior user behavior can identify a user and their habits. Because of the current and growing algorithmic and computational capabilities of adversaries, provable privacy guarantees as a function of the degree of anonymization and obfuscation of the traces are necessary. Our previous work has established the requirements on anonymization and obfuscation in the case that data traces are independent between users. However, the data traces of different users will be dependent in many applications, and an adversary can potentially exploit such. In this paper, we consider the impact of correlation between user traces on their privacy. First, we demonstrate that the adversary can readily identify the association graph, revealing which user data traces are correlated. Next, we demonstrate that the adversary can use this association graph to break user privacy with significantly shorter traces than in the case when traces are independent between users, and that independent obfuscation of the data traces is often insufficient to remedy such. Finally, we discuss how the users can employ dependency in their obfuscation to improve their privacy.**

***Index Terms*—Internet of Things (IoT), Privacy-Protection Mechanisms (PPM), Information Theoretic Privacy, Anonymization, Obfuscation, Inter-User Correlation.**

I. Introduction

Many modern applications exploit a user's characteristics, both their past choices and present state, to enhance user experience. For example, emerging Internet of Things (IoT) applications include smart homes, health care, and connected vehicles that will smartly tune their response to a given user, such as a connected vehicle application optimizing a route based on the current location of the vehicle and traffic conditions. Such applications require that a user provide potentially sensitive data; hence, questions arise about how much user privacy is compromised. Even if user data traces are anonymized, statistical matching of the current data trace with prior user behavior can identify the user and their characteristics [1], [2]. And studies have indicated that privacy concerns could significantly hamper the penetration of IoT applications [3]–[5].

Our previous work has introduced the notion of "perfect privacy" [6]. In particular, with rapid advances in algorithms and computation, information-theoretic guarantees that demonstrate that sensitive information does not leak to a powerful adversary are critical. The work of [6]–[8] considered the degree of user anonymization and data obfuscation required to obtain perfect privacy in the case when the data traces of different users are independent of one another. In that work, we have considered the case of independent and identically distributed (i.i.d.) samples from a given user and the case when there is temporal correlation within the trace of a given user [7], [8], as have others under different metrics [9]–[12].

There are many applications where there is correlation between the traces of different users. For example, friends tend to travel together or might meet at given places, hence introducing dependency between locations. However, there is relatively limited work in this area [13]–[17], particularly from a fundamental perspective. Hence, here we investigate what the metrics of [6]–[8] require in terms of user anonymization and data obfuscation to preserve privacy in the case of correlation between the data traces of different users.

We model dependence between user traces with an association graph, where an edge between the vertices corresponding to a pair of users indicates dependency between their data traces. We first demonstrate that the adversary can readily determine this association graph. Armed with this association graph, the adversary can attempt to identify the users, and we show that this provides the adversary with a significant advantage versus the case when the data traces of different users are independent of one another. This suggests that, unless additional countermeasures are employed, the results of [6]–[8], [18], [19] for independent traces are overly optimistic when user traces are correlated. We next consider countermeasures. First, we demonstrate that adding independent obfuscation to user data samples is often ineffective in improving the users' privacy. Finally, we demonstrate that, if users with correlated traces can jointly design their obfuscation, user privacy can be significantly improved.

More references, additional discussion, and proofs of main results are provided in the long version of the paper [20].

This work was supported by National Science Foundation under grants CCF–1421957 and CNS-1739462.

978-1-5386-4780-6/18/$31.00©2018 IEEE

II. Framework

We employ a similar framework to [6], [8]. The system has n users, and $X_u(k)$ is the sample of the data of user u at time k. Our main goal is protecting $X_u(k)$ from a strong adversary ($\mathcal{A}$) who has full knowledge of the (unique) marginal probability distribution function of the data samples for each user based on previous observations or other resources. In order to achieve data privacy of users, both anonymization and obfuscation techniques can be used as shown in Figure 1. In Figure 1, $Z_u(k)$ shows the (reported) sample of the data of user u at time k after applying obfuscation, and $Y_u(k)$ shows the (reported) sample of the data of user u at time k after applying anonymization. Let $m = m(n)$ be the number of data points after which the pseudonyms of users are changed in the anonymization. To break the anonymization, the adversary tries to estimate $X_u(k)$, $k = 1, 2, \cdots, m$, from m observations per user by matching the sequence of observations to the known statistical characteristics of the users.

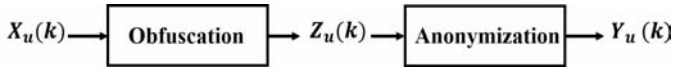

Fig. 1: Notation for sequences after applying obfuscation and anonymization to users' data samples.

Let $\mathbf{X}_u$ be the $m \times 1$ vector containing the samples of the data of user u, and $\mathbf{X}$ be the $m \times n$ matrix with u^{th} column equal to $\mathbf{X}_u$;

$$\mathbf{X}_u = \left[X_u(1), X_u(2), \cdots, X_u(m)\right]^T$$

$$\mathbf{X} = \left[\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n\right].$$

Data Samples Model: We assume users' data samples can have r possibilities $(0, 1, \cdots, r-1)$. Thus, according to a user-specific probability distribution, $X_u(k)$ is equal to a value in $\{0, 1, \cdots, r-1\}$ at any time, and, per above, these user-specific probability distributions are known to the adversary ($\mathcal{A}$) and form the basis upon which he performs (statistical) matching.

Association Graph: An association graph or dependency graph is an undirected graph representing dependencies of users with each other. Let $G(V, F)$ denote the association graph with set of nodes V, ($|V| = n$), and set of edges F. Two vertices (users) are connected if their data sets are dependent. More specifically,

- $(u, u') \in F$ if $I(X_u(k); X_{u'}(k)) > 0$,
- $(u, u') \notin F$ if $I(X_u(k); X_{u'}(k)) = 0$,

where $I(X_u(k); X_{u'}(k))$ is the mutual information between the k^{th} data sample of user u and user u'.

Obfuscation Model: Obfuscation perturbs the users' data samples. Each user has only limited knowledge of the characteristics of the overall population, so usually a simple distributed method in which the samples of the data of each user are reported with error with a certain probability is employed. Note that this probability itself is generated randomly for each user. Let $\mathbf{Z}_u$ be the vector which contains the obfuscated versions of user u's data sample, and $\mathbf{Z}$ be the collection of $\mathbf{Z}_u$ for all users,

Anonymization Model: In the anonymization technique, the identity of the users is perturbed. Anonymization is modeled by a random permutation Π on the set of n users. Let $\mathbf{Y}_u$ be the vector which contains the anonymized version of $\mathbf{Z}_u$ (the obfuscated version of data of user u), and $\mathbf{Y}$ is the collection of $\mathbf{Y}_u$ for all users, thus $\mathbf{Y}_u = \mathbf{Y}_{\Pi^{-1}(u)}$ and $\mathbf{Y}_{\Pi(u)} = \mathbf{Y}_u$.

Adversary Model: We assume the adversary has full knowledge of the marginal probability distribution function of each of the users on $\{0, 1, \ldots, r-1\}$. The adversary also knows the anonymization mechanism, but does not know the realization of the random permutation. The adversary knows the obfuscation mechanism, but does not know the realization of the noise parameters. Also, the adversary knows the association graph $G(V, F)$, but does not necessarily know the exact nature of dependency. That is, while the adversary knows the marginal distributions $X_u(k)$ as well as which pairs of users have strictly positive mutual information, he might not know the joint distributions or even the values of mutual informations $I(X_u(k); X_{u'}(k))$. It is critical to note that we assume the adversary does not have any auxiliary information or side information about users' data.

Definition 1. User u has *perfect privacy* [6] if and only if

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

where $I\left(X_u(k); \mathbf{Y}\right)$ denotes the mutual information between a sample of the data of user u at time k and the collection of the adversary's observations for all of the users.

Definition 2. User u has *no privacy* [8] if and only if there exists an algorithm for the adversary to estimate $X_u(k)$ perfectly as n goes to infinity. In other words, as $n \to \infty$,

$$\forall k \in \mathbb{N}, \quad P_e(u, k) \triangleq P\left(\widetilde{X_u(k)} \neq X_u(k)\right) \to 0,$$

where $\widetilde{X_u(k)}$ is the estimated value of $X_u(k)$ by the adversary.

III. Impact of Correlation Between Users on Privacy Using Anonymization

In this section, we consider only anonymization and thus the obfuscation block in Figure 1 is not present.

A. *i.i.d. Two-State Model*

There is potentially correlation between the data of different users, but we assume here that the sequence of data for any individual user is *i.i.d.*. The *i.i.d.* model would apply directly to data that is sampled at a low rate. In addition, understanding the i.i.d. case can also be considered the first step toward understanding the more complicated case where there is dependency.

We first consider the *i.i.d.* two-state ($r = 2$) case, where the sample of the data of user u at any time is a Bernoulli random variable with parameter p_u, which we define as the probability of user u being at state 1. Thus,

$$X_u(k) \sim Bernoulli\left(p_u\right).$$

The parameters p_u, $u = 1, 2, \cdots, n$ are drawn independently from a continuous density function, $f_P(p_u)$, on the $(0, 1)$ interval. The density $f_P(p_u)$ might be unknown, so all that is assumed here is that such a density exists. From the results of the paper, it will be evident that knowing or not knowing $f_P(p_u)$ does not change the results asymptotically. Further, we assume there are $\delta_1, \delta_2 > 0$ such that[1]:

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

The adversary knows the values of p_u, $u = 1, 2, \cdots, n$, and uses this knowledge to match the observed traces to the users. We will use capital letters (i.e., P_u) when we are referring to the random variable, and use lower case (i.e., p_u) to refer to the realization of P_u.

A vector containing the permutation of those probabilities after anonymization is $\widetilde{\mathbf{P}}$, where $\widetilde{P}_u = P_{\Pi^{-1}(u)}$ and $\widetilde{P}_{\Pi(u)} = P_u$. As a result, for $u = 1, 2, ..., n$, the distribution of the data symbols for the user with pseudonym u is given by:

$$Y_u(k) \sim Bernoulli\left(\widetilde{P}_u\right) \sim Bernoulli\left(P_{\Pi^{-1}(u)}\right).$$

For this case, dependency and correlation of the data samples are equivalent, that is, we can say:

- $(u, u') \in F$ if $\rho(X_u(k), X_{u'}(k)) = \rho_{uu'} > 0$,
- $(u, u') \notin F$ if $\rho(X_u(k), X_{u'}(k)) = \rho_{uu'} = 0$,

where $\rho_{uu'}$ is the correlation coefficient between the data of user u and that of user u'. The adversary knows the association graph $G(V, F)$, but does not necessarily know the correlation coefficient $(\rho_{uu'})$ for each specific $(u, u') \in F$.

Critical to compromising the privacy of the users will be the adversary's ability to match empirical correlation properties of the data traces to the known structure of the (ensemble) correlation between users. First, we show that the adversary can reliably reconstruct the entire association graph for *the anonymized version of the data* (i.e. the observed data traces) with relatively few observations.

Lemma 1. Consider a general association graph $G(V, F)$. If the adversary obtains $m = (\log n)^3$ anonymized observations per user, he/she can construct $\widetilde{G} = \widetilde{G}(\widetilde{V}, \widetilde{F})$, where $\widetilde{V} = \{\Pi(u) : u \in V\} = V$, such that for all $u, u' \in V$; $(u, u') \in F$ iff $\left(\Pi(u), \Pi(u')\right) \in \widetilde{F}$. We write this statement as $P(\widetilde{F} = F) \to 1$.

The structure of the association graph (G) can leak a significant amount of information. For example, in Figure 2, the identity map is the only automorphism of the association graph G. Thus, it is obvious that the adversary can uniquely identify all of the users if he/she can reconstruct the association graph.

To be able to derive further results, we need to make some assumptions on the structure of the association graph. For the rest of the paper, we consider a graph structure shown

[1]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.

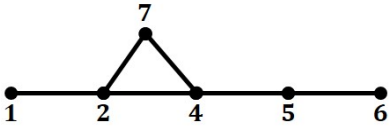

Fig. 2: One example of an association graph for which the identity map is the only automorphism.

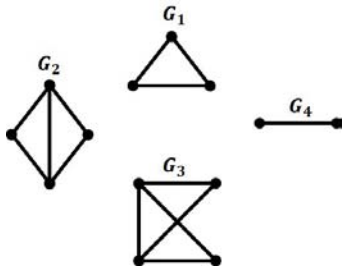

Fig. 3: Association graph consists of some disjoint subgraphs (G_j), where G_j is a connected graph on s_j vertices.

in Figure 3, where the association graph consists of f disjoint subgraphs,

$$G = G_1 \cup G_2 \cup \cdots \cup G_f,$$

where subgraph G_j is a connected graph on s_j vertices. In particular, each subgraph G_j refers to a group of "friends" or "associates" such that their data sets are dependent, and we will denote its association graph as $G_j(V_j, F_j)$, where $|V_j| = s_j$.

The following theorem states that if the number of observations per user (m) is significantly larger than $n^{\frac{2}{s}}$ in this two-state model, then the adversary can successfully de-anonymize the users in any group of size s.

Theorem 1. For the above two-state model, if $\mathbf{Y}$ is the anonymized version of $\mathbf{X}$ as defined above, the size of a group is s, and $m = cn^{\frac{2}{s}+\alpha}$, where $\alpha > 0$, then user 1 has no privacy as n goes to infinity.

Discussion: It is insightful to compare this result to Theorem 1 in [6], where it is stated that if the users are not correlated, then all users have perfect privacy as long as the number of adversary's observations per user (m) is smaller than $O(n^2)$. Here, Theorem 1 states that with much smaller m the adversary can de-anonymize all users. Therefore, we see that correlation can significantly reduce the privacy of users.

B. i.i.d. r-States Model

Now, assume users' data samples can have r possibilities $(0, 1, \cdots, r-1)$, and $p_u(i)$ gives the probability of user u having data sample i. We define the vector $\mathbf{p}_u$ as

$$\mathbf{p}_u = [p_u(1), p_u(2), \cdots, p_u(r-1)]^T, \quad \mathbf{P} = \left[\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n\right].$$

We also assume $\mathbf{p}_u$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, which has support on a

subset of the $(0,1)^{r-1}$ hypercube. In particular, define the range of the distribution as

$$\mathcal{R}_{\mathbf{P}} = \{(x_1, x_2, \cdots, x_{r-1}) \in (0,1)^{r-1} : \\ x_i > 0, x_1 + x_2 + \cdots + x_{r-1} < 1\}.$$

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

$$\begin{cases} \delta_1 \le f_{\mathbf{P}}(\mathbf{p}_u) \le \delta_2, & \mathbf{p}_u \in \mathcal{R}_{\mathbf{P}}. \\ f_{\mathbf{P}}(\mathbf{p}_u) = 0, & \mathbf{p}_u \notin \mathcal{R}_{\mathbf{P}}. \end{cases}$$

Theorem 2. For the above r-states model, if $\mathbf{Y}$ is the anonymized version of $\mathbf{X}$ as defined above, the size of a group is s, and $m = cn^{\frac{2}{(r-1)s}+\alpha}$, where $\alpha > 0$, then user 1 has no privacy as n goes to infinity.

C. Markov Chain Model

In Sections III-A and III-B, we assumed each user's data patterns was $i.i.d.$; however, in this section users' data patterns are modeled using Markov chains in which each user's data samples are dependent over time. In this model, we again assume there are r possibilities for each data point, i.e., $X_u(k) \in \{0, 1, \cdots, r-1\}$.

More specifically, each user's data is modeled by a Markov chain with r states. It is assumed that the Markov chains of all users have the same structure, but have different transition probabilities. Let E be the set of edges in the assumed transition graph, so $(i,l) \in E$ if there exists an edge from state i to state l, meaning that $p_u(i,l) = P\left(X_u(k+1) = l | X_u(k) = i\right) > 0$. The transition matrix is a square matrix used to describe the transitions of a Markov chain; thus, different users can have different transition probability matrices. Note for each state i, we have $\sum_{l=1}^{r} p_u(i,l) = 1$, so the adversary can focus on a subset of size $d = |E| - r$ of transition probabilities for recovering the entire transition matrix. So we have

$$\mathbf{p}_u = [p_u(1), p_u(2), \cdots, p_u(d)]^T, \quad \mathbf{P} = \left[\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_n\right].$$

We also consider $p_u(i)$'s are drawn independently from some continuous density function, $f_P(\mathbf{p}_u)$, on the $(0,1)^{|E|-r}$ hypercube. Define the range of the distribution as

$$\mathcal{R}_{\mathbf{P}} = \left\{(x_1, \cdots, x_d) \in (0,1)^d : x_i > 0, x_1 + x_2 + \cdots + x_d < 1\right\}.$$

As before, we assume there are $\delta_1, \delta_2 > 0$, such that

$$\begin{cases} \delta_1 \le f_{\mathbf{P}}(\mathbf{p}) \le \delta_2, & \mathbf{p} \in \mathcal{R}_{\mathbf{p}} \\ f_{\mathbf{P}}(\mathbf{p}) = 0, & \mathbf{p} \notin \mathcal{R}_{\mathbf{p}} \end{cases}$$

Theorem 3. For an irreducible, aperiodic Markov chain model, if $\mathbf{Y}$ is the anonymized version of $\mathbf{X}$ as defined above, the size of a group is s, and $m = cn^{\frac{2}{s(|E|-r)}+\alpha}$, where $\alpha > 0$, then user 1 has no privacy as n goes to infinity.

IV. Privacy using Anonymization and Obfuscation

Here we consider the case when both anonymization and obfuscation techniques are employed, as shown in Figure 1. We assume similar obfuscation to [8].

A. $i.i.d.$ Two-State Model

Again, let us start with the $i.i.d.$ two-state model. 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.

To obfuscate the data samples, for each user u we independently generate a random variable R_u that is uniformly distributed between 0 and a_n. The value of R_u shows the probability that the user's data sample is changed to a different value by obfuscation, and a_n is termed the "noise level" of the system.

The effect of the obfuscation is to alter the probability distribution function of each user's data samples in a way that is unknown to the adversary, since it is independent of all past activity of the user, and hence the obfuscation inhibits user identification. For each user, R_u is generated once and is kept constant for the collection of samples of length m, thus providing a very low-weight obfuscation algorithm.

The $Z_u(k)$'s are $i.i.d.$ with a Bernoulli distribution; thus,

$$Y_u(k) \sim Bernoulli\left(\widehat{p}_u\right),$$

where $\widehat{p}_u$ is is the probability that an obfuscated data sample of user u is equal to one, so

$$\begin{aligned} \widehat{p}_u &= p_u(1 - R_u) + (1 - p_u)R_u \\ &= p_u + \left(1 - 2p_u\right) R_u. \end{aligned}$$

Define the vector $\widehat{\mathbf{P}}$, which contains the obfuscated probabilities:

$$\widehat{\mathbf{P}} = \left[\widehat{p}_1, \widehat{p}_2, \cdots, \widehat{p}_n\right],$$

and the vector containing the permutation of those probabilities after anonymization as $\widetilde{\mathbf{P}}$. As a result, for $u = 1, 2, ..., n$,

$$Y_u(k) \sim Bernoulli\left(\widetilde{p}_u\right).$$

Theorem 4. For the above two-state model, if $\mathbf{Z}$ is the obfuscated version of $\mathbf{X}$, and $\mathbf{Y}$ is the anonymized version of $\mathbf{Z}$ as defined above, the size of a group is s, and

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

then user 1 has no privacy as n goes to infinity.

Discussion: It is insightful to compare this result to Theorem 1 in [8], which stated that if the users are not correlated, then all users have perfect privacy as long as the number of the adversary's observations per user (m) is smaller than $O(n^2)$ or the noise level (a_n) used to obfuscate the users' data samples is larger than $O(n^{-1})$. Here, Theorem 4 states that with much smaller m or much larger a_n the adversary can de-anonymize and de-obfuscate all users with vanishing error probability. Therefore, we see that correlation can significantly reduce the privacy of users.

B. $i.i.d.$ r-States Model and Markov Chain Model

Similar to sections III-B and III-C, we can extend the results of Section IV-A to the r-states models well as the Markov chain models. The details are provided in the extended version [20].

V. Achieving Perfect Privacy in the Presence of Correlation

Here, we discuss how we can improve privacy in the presence of correlation. First note that independent obfuscation alone is not sufficient even at a high noise level, because it cannot change the association graph. Thus, we suggest that associated users collaborate together to increase their privacy. For clarity, we focus on the two-state model with $s_j \leq 2$; thus, there are some users that are connected together and there are also some isolated users. The asymptotic noise level is defined as the highest probable percentage of data points that are corrupted [20] .

Theorem 5. For the two-state model, if $\mathbf{Z}$ is the obfuscated version of $\mathbf{X}$, $\mathbf{Y}$ is the anonymized version of $\mathbf{Z}$, and the size of all subgraphs are less than or equal to 2, there exists an anonymization/obfuscation scheme such that for all $(u, u') \in F$, the asymptotic noise level for users u and u' is at most

$$a(u,u') = \frac{\left|\mathrm{Cov}(X_u(k), X_{u'}(k))\right|}{\max\{p_u, p_{u'}, 1-p_u, 1-p_{u'}\}},$$

to achieve perfect privacy for all users.

VI. Conclusion

A sophisticated adversary can threaten user privacy by employing statistical matching of user data traces to prior behavior. Our previous work has considered the requirements on anonymization and obfuscation for "perfect" user privacy to be maintained when traces are independent between users. But traces are rarely independent, as relationships between users establish dependence in their behavior. We have shown here that such dependence can have a significant impact on user privacy, as the anonymization employed must be significantly increased to preserve perfect privacy, and often no degree of independent obfuscation of the traces can be effective. We also present preliminary results on dependent obfuscation to preserve user privacy.

References

[1] F. M. 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.

[2] 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, 2014.

[3] N. Apthorpe, D. Reisman, and N. Feamster, "A Smart Home is No Castle: Privacy Vulnerabilities of Encrypted IoT Traffic," in *Workshop on Data and Algorithmic Transparency*, 2016.

[4] P. Porambag, M. Ylianttila, C. Schmitt, P. Kumar, A. Gurtov, and A. V. Vasilakos, "The quest for privacy in the internet of things," *IEEE Cloud Computing*, vol. 3, no. 2, pp. 36–45, 2016.

[5] A. Ukil, S. Bandyopadhyay, and A. Pal, "IoT-privacy: To be private or not to be private," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Toronto, ON, Canada: IEEE, 2014, pp. 123–124.

[6] Z. Montazeri, A. Houmansadr, and H. Pishro-Nik, "Achieving Perfect Location Privacy in Wireless Devices Using Anonymization," *IEEE Transaction on Information Forensics and Security*, vol. 12, no. 11, pp. 2683–2698, 2017.

[7] N. Takbiri, A. Houmansadr, D. Goeckel, and H. Pishro-Nik, "Matching anonymized and obfuscated time series to users' profiles," https://arxiv.org/abs/1710.00197, May 2018, [Online; accessed 02-May-2018].

[8] N. Takbiri, A. Houmansadr, D. L. Goeckel, and H. Pishro-Nik, "Limits of location privacy under anonymization and obfuscation," in *International Symposium on Information Theory (ISIT)*. Aachen, Germany: IEEE, 2017, pp. 764–768.

[9] R. Al-Dhubhani and J. Cazalas, "Correlation analysis for geo-indistinguishability based continuous lbs queries," in *2nd International Conference on Anti-Cyber Crimes (ICACC)*. Abha, Saudi Arabia: IEEE, 2017.

[10] S. Zhang, Q. Ma, T. Zhu, K. Liu, L. Zhang, W. He, and Y. Liu, "Plp: Protecting location privacy against correlation-analysis attack in crowdsensing," in *44th International Conference on Parallel Processing (ICPP)*. Beijing, China: IEEE, 2015.

[11] H. Liu, X. Li, and H. Li, "Spatiotemporal correlation-aware dummy-based privacy protection scheme for location-based services," in *IEEE Conference on Computer Communications (INFOCOM)*. Atlanta, GA, USA: IEEE, 2017.

[12] Y. Xiao and L. Xiong, "Protecting locations with differential privacy under temporal correlations," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1298–1309.

[13] D. Kifer and A. Machanavajjhala, "No free lunch in data privacy," in *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. Athens, Greece: ACM, 2011, pp. 193–204.

[14] T. Zhu, P. Xiong, G. Li, and W. Zhou, "Correlated differential privacy: Hiding information in non-iid data set," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 229–242, 2015.

[15] L. Ou, Z. Qin, Y. Liu, H. Yin, Y. Hu, and H. Chen, "Multi-user location correlation protection with differential privacy," in *IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE], year=2016, address= Wuhan, China.

[16] B. Yang, I. Sato, and H. Nakagawa, "Bayesian differential privacy on correlated data," in *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. Melbourne, Victoria, Australia: ACM, 2015, pp. 747–762.

[17] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, , and X. Xiao, "Privbayes: Private data release via bayesian networks," in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, pp. 1423–1434.

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

[19] N. Takbiri, A. Houmansadr, D. Goeckel, and H. Pishro-Nik, "Statistical matching in the presence of anonymization and obfuscation: Non-asymptotic results in the discrete case," in *Annual Conference on Information Science and Systems (CISS)*. Princeton,NJ, USA: IEEE, 2018.

[20] N. Takbiri, A. Houmansadr, D. Goeckel, and H. Pishro-Nik, "Privacy against statistical matching: Inter-user correlation," http://www.ecs.umass.edu/ece/pishro/Papers/correlation.pdf, January 2018, [Online; accessed 07-April-2018].