Differential Privacy-Preserving User Linkage across Online Social Networks

Xin Yao*‡, Rui Zhang†, Yanchao Zhang‡
*School of Computer Science and Engineering, Central South University, Changsha, Hunan, 410082, China
†Department of Computer and Information Sciences, University of Delaware, Newark, DE 19716, USA
‡School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA

***Abstract*—Many people maintain accounts at multiple online social networks (OSNs). Multi-OSN user linkage seeks to link the same person's web profiles and integrate his/her data across different OSNs. It has been widely recognized as the key enabler for many important network applications. User linkage is unfortunately accompanied by growing privacy concerns about real identity leakage and the disclosure of sensitive user attributes. This paper initiates the study on privacy-preserving user linkage across multiple OSNs. We consider a social data collector (SDC) which collects perturbed user data from multiple OSNs and then performs user linkage for commercial data applications. To ensure strong user privacy, we introduce two novel differential privacy notions, ϵ-attribute indistinguishability and ϵ-profile indistinguishability, which ensure that any two users' similar attributes and profiles cannot be distinguished after perturbation. We then present a novel Multivariate Laplace Mechanism (MLM) to achieve ϵ-attribute indistinguishability and ϵ-profile indistinguishability. We finally propose a novel differential privacy-preserving user linkage framework in which the SDC trains a classifier for user linkage across different OSNs. Extensive experimental studies based on three real datasets confirm the efficacy of our proposed framework.**

***Index Terms*—User linkage, differential privacy, online social networks.**

I. Introduction

The past decade has witnessed the rise of Online Social Networks (OSNs). Given various OSNs with different features, it is common for people to register and use multiple accounts at different OSNs. For example, many people publish their emotions and opinions about what they see and hear on Twitter, write restaurant reviews on Yelp, and find career opportunities on LinkedIn. Statistics released by Brandwatch show that every OSN user has 7.6 OSN accounts on average.[1] Besides, a recent survey shows that about 73% of OSN users have accounts at more than one OSNs.[2]

User linkage—also known as user recognition, anchor linking, user resolution, etc.—has been an important issue due to so many multi-OSN users. The primary goal of user linkage is to identify the same user and link his/her web profiles across different OSNs. User linkage has been widely recognized as the key enabler for various social network applications. For example, it can provide better understanding of users' interests and behaviors [1]; it can also mitigate the cold-start and data sparsity problems of social recommendation [2] and prediction systems [3]. User linkage has attracted growing attention from both the academia and industry. For instance, Goga *et al.* [4] recently reported a system that can accurately find identical identities of about 30% of OSN users.

User linkage is challenged by privacy concerns about real identity leakage and the disclosure of sensitive user attributes [5]–[7]. In particular, attackers can infer the real identity of an OSN user more accurately by jointly considering his/her profiles and activities across multiple OSNs. Similarly, attackers can learn additional information about a user by examining his/her social activities across multiple OSNs.

This paper makes the first attempt to study **privacy-preserving user linkage** across multiple OSNs. Specifically, we consider a *social data collector* (SDC) introduced in [8], which collects user data from multiple OSNs and intends to link users across different OSNs. To protect user privacy, each OSN perturbs its user data before sharing them with the SDC. Given perturbed user data, the SDC predicts whether a given pair of users on different OSNs are associated with the same real person. After performing user linkage, the SDC can aggregate the multi-OSN data of the same users and resell such linked user data to end data consumers to facilitate various network applications. One such application is privacy-preserving recommendation and prediction systems like [9], which take perturbed OSN user data as input and output the recommendation or prediction results.

Privacy-preserving user linkage across multiple OSNs poses unique challenges. Traditional countermeasures against identity leakage and attribute disclosure include k-anonymity [10], ℓ-diversity [11], t-closeness [12], etc., which can offer satisfactory privacy protection for well-defined relational data but lack a rigorous theoretical guarantee. Recently, Local Differential Privacy (LDP) [13] has been widely recognized as a strong and mathematically rigorous privacy-preserving framework and has found success in many statistic problems [14]–[23]. However, if directly applied to privacy-preserving user linkage, LDP suffers from a major limitation. In particular, since OSN user attributes often have large value ranges, satisfying LDP would require injecting large noise and then result in low data utility and also low user-linkage accuracy. How to achieve high linkage prediction accuracy while protecting user privacy remains an open challenge.

[1]https://www.brandwatch.com/blog/amazing-social-media-statistics-and-facts/

[2]https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/

978-0-7381-3207-5/21/$31.00 ©2021 IEEE

978-1-6654-1494-4/21/$31.00 ©2021 IEEE

We propose a novel framework for differential privacy-preserving user linkage across multiple OSNs. In our framework, each OSN perturbs its user data with differential privacy techniques and then shares perturbed data with a semi-trusted SDC. Given such perturbed user data, the SDC uses a proper machine learning technique to train a classification model to predict whether any two users from different OSNs correspond to the same real person. To address the limitation of LDP, we introduce two new privacy notions, *ϵ-attribute indistinguishability* and *ϵ-profile indistinguishability*. Different from standard LDP that aims to ensure any two users are indistinguishable after perturbation, ϵ-attribute indistinguishability and ϵ-profile indistinguishability ensure that any two users with similar attributes and profiles are indistinguishable by taking the distance between their attributes and profiles into account, respectively. The smaller the distance between two user attributes (or profiles), the more indistinguishable they are after perturbation, and vice versa. Relaxing the privacy protection from LDP to ϵ-profile indistinguishability can effectively reduce the amount of injected noise to improve data utility and user-linkage accuracy. We further propose a novel Multivariate Laplace Mechanism (MLM) which adds correlated noise to multi-dimensional user attributes and satisfies ϵ-attribute indistinguishability.

We evaluate the proposed framework and MLM with three public real datasets from `Twitter`, `MySpace`, and `Last.fm`, which include 28,199, 9,993, and 7,661 users, respectively. Our results show that the user-linkage accuracy under MLM is significantly higher than existing perturbation mechanisms and improves much faster as the privacy budget ϵ increases. For example, when $\epsilon = 0.5$, the accuracy score of Logistic Regression on the Twitter-MySpace dataset pair with MLM reaches 63.55%, while those with other mechanisms are only about 50%. Moreover, our results confirm the advantages of adding correlated noise to different user attributes.

The rest of this paper is structured as follows. Section II introduces the system model, adversary model, and problem formulation. Section III briefs some background on differential privacy. Section IV presents the proposed framework. Section V conducts the privacy analysis. Section VI presents the evaluation results. Section VII discusses the related work. Section VIII concludes this paper.

II. Problem Formulation and Adversary Model

A. Problem Formulation

We consider an OSN user-linkage system comprising one SDC and two OSNs denoted by U and V, respectively. Our system can easily support more OSNs, of which user linkage operations are individually performed for each pair. Each OSN shares its user data to the SDC without violating its privacy guarantees to its users. The SDC employs machine-learning techniques to train a classifier for linking users on U and V to the same real person. Finally, the SDC can sell linked user data to end data consumers for various important OSN applications such as recommendation and prediction.

While OSN users socialize with each other in various ways, their data can commonly be classified into personal attributes and social activity contents. Personal attributes can be in different data formats. For example, age, weight, height, and income are usually numeric values, while home/work address and recent activities are usually text strings. To facilitate subsequent perturbation and profile linkage, we convert every text string into a numeric value. Specifically, we first decompose each text string into a set of substrings via the standard n-gram technique [24]. We then convert the resulting substring set into a numerical value using SimHash [25]. The locality sensitivity of SimHash indicates that two similar text strings must have similar SimHash values, and vice versa. There could also be user attributes in the form of categorical data such as gender and race. We focus on numeric and text data in this paper and leave the support for categorical data as our future work.

We use the following notations. OSN U consists of $n \gg 1$ users denoted by $U = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$. Each $\boldsymbol{u}_i \in U$ is represented by a γ-dimension attribute vector $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,\gamma})$, where γ is the number of attributes, and each $u_{i,k}$ ($1 \le i \le n$, $1 \le k \le \gamma$) is the kth attribute in $[0, 1]$ after proper normalization. Similarly, we assume that OSN V is composed of $m \gg 1$ users denoted by $V = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, in which each user $\boldsymbol{v}_j \in V$ is represented by an attribute vector $\boldsymbol{v}_j = (v_{j,1}, \ldots, v_{j,\gamma})$ in which each attribute value also takes value in [0,1] after proper normalization.

We seek to develop an effective framework for differential privacy-preserving user linkage across U and V. In our framework, U and V each employ a differential privacy mechanism $\mathcal{M}$ to perturb each $\boldsymbol{u}_i \in U$ to obtain $\widetilde{\boldsymbol{u}}_i$ and each $\boldsymbol{v}_j \in V$ to obtain $\widetilde{\boldsymbol{v}}_j$, respectively. Given two perturbed user datasets $\widetilde{U} = \{\widetilde{\boldsymbol{u}}_i\}_{i=1}^n$ and $\widetilde{V} = \{\widetilde{\boldsymbol{v}}_j\}_{j=1}^m$, the SDC aims to determine whether each user pair $(\widetilde{\boldsymbol{u}}_i, \widetilde{\boldsymbol{v}}_j)$ is associated with the same real person for all $1 \le i \le n$ and $1 \le j \le m$. The detailed requirements for $\mathcal{M}$ are postponed to Section III.

B. Adversary Model

We assume that each OSN is an independent, trusted entity and has legal obligations to protect user privacy. As a standard practice, each OSN replaces every user ID with a distinct anonymous ID before sharing its user dataset with the SDC. In addition, different OSNs do not share their user data with each other due to business competition and lack of trust.

The SDC is assumed to be honest-but-curious. Specifically, the SDC honestly performs user linkage operations. Following the prior work [8], the SDC is curious in the sense that it attempts to link selected anonymous IDs in the received datasets to real IDs on the OSN platforms whereby to learn additional social activities and other sensitive information of the victims from available side information [26].

III. Preliminaries

In this section, we introduce local differential privacy pertaining to the proposed two privacy notions. Local differential privacy [27] is a powerful technique to ensure data privacy against a curious data collector, which is defined as follows.

Definition 1. *A randomized mechanism $\mathcal{M}$ satisfies ϵ-local differential privacy if and only if*

$$\frac{\Pr[\mathcal{M}(u) = O]}{\Pr[\mathcal{M}(u') = O]} \leq e^{\epsilon},$$

for any two inputs u, u' and for any possible output $O \in \mathsf{range}(\mathcal{M})$.

Different from centralized differential privacy for which perturbation is performed by the data collector, each OSN perturbs its user data locally under LDP. The parameter ϵ is commonly referred to as the privacy budget and can be used to measure the achievable privacy of $\mathcal{M}$. The smaller the ϵ is, the higher level of privacy $\mathcal{M}$ provides, and vice versa.

The Laplace mechanism is a classical technique that achieves LDP by adding the Laplace noise to user attributes. Let $\triangle u_k$ be the size of attribute k's data range for all $1 \leq k \leq \gamma$. Given an arbitrary user-attribute vector $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,\gamma})$, the Laplace mechanism outputs the perturbed one as

$$\widetilde{\boldsymbol{u}}_i = (u_{i,1} + \mathsf{Lap}(\frac{\triangle u_{i,1}}{\epsilon_1}), \ldots, u_{i,\gamma} + \mathsf{Lap}(\frac{\triangle u_{i,\gamma}}{\epsilon_\gamma})) \ ,$$

where $\mathsf{Lap}(\lambda)$ denotes a random sample generated from a Laplace distribution of scale λ with the probability density function

$$f(x) = \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda}) \ .$$

The following theorem shows the privacy guarantee offered by the Laplace mechanism.

Theorem 1. *The Laplace mechanism $\mathcal{M}$ satisfies $\epsilon_{\max}$-local differential privacy [28], where*

$$\epsilon_{\max} = \max(\epsilon_1, \ldots, \epsilon_\gamma) \ .$$

IV. Differential Privacy-Preserving User Linkage

In this section, we first introduce two novel differential privacy notions and then present a novel multivariate Laplace mechanism along with its analysis. Finally, we introduce our differential privacy-preserving user linkage framework.

A. ϵ-Profile/Attribute Indistinguishability

While LDP is a classical notion for privacy protection, directly applying it to multi-OSN user linkage suffers from one major limitation. Specifically, satisfying LDP requires that for any two user-attribute vectors $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$, the ratio of their probabilities of being transformed into any perturbed attribute vector is upper-bounded by e^{ϵ}. In other words, LDP offers the same level of privacy protection to two highly dissimilar users as to two highly similar users. Since OSN user attributes commonly have large value ranges, satisfying LDP would require injecting large noise to accommodate the most diverse pair of user-attribute vectors, which results in low data utility and low user-linkage accuracy.

Inspired by the notions of ϵ-geo-indistinguishability [29] and ϵ-text indistinguishability [8], we introduce two new privacy notions in the context of user linkage, *ϵ-attribute indistinguishability* and *ϵ-profile indistinguishability*. Instead of ensuring that any two user-attribute vectors are indistinguishable after perturbation, we aim to ensure that similar user-attribute vectors are indistinguishable after perturbation, and the level of privacy protection depends on the similarity or distance between two attributes or profiles. More specifically, ϵ-attribute indistinguishability ensures that two users' attributes, e.g., age, are indistinguishable after perturbation. The smaller the distance between the two users' attributes, the more indistinguishable they are, and vice versa. Similarly, ϵ-profile indistinguishability ensures two users with similar attribute vectors are not distinguishable after perturbation.

For any user-attribute vector $\boldsymbol{u}_i \in U$, we denote by $\mathcal{M}(\boldsymbol{u}_i)$ the perturbed output by the mechanism $\mathcal{M}$. We also abuse the notation to let $\mathcal{M}(u_{i,k})$ denote the kth attribute of $\mathcal{M}(\boldsymbol{u}_i)$. We give two definitions of ϵ-attribute indistinguishability and ϵ-profile indistinguishability below.

Definition 2. *(ϵ-**attribute indistinguishability**) Given a set of user-attribute vectors $U = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$, a mechanism $\mathcal{M}(\cdot)$ satisfies ϵ-attribute indistinguishability with respect to the kth attribute if and only if ,*

$$\frac{\Pr[\mathcal{M}(u_{i,k}) = t]}{\Pr[\mathcal{M}(u_{j,k}) = t]} \leq e^{\epsilon |u_{i,k} - u_{j,k}|}, \tag{1}$$

for any pair of attributes $u_{i,k}$ and $u_{j,k}$ and any possible perturbed attribute value $t \in \mathbb{R}$.

The notion of ϵ-attribute indistinguishability indicates that the ratio of the probabilities that any two user attributes $u_{i,k}$ and $u_{j,k}$ are transformed by mechanism $\mathcal{M}(\cdot)$ into the same value is upper bounded by $e^{\epsilon |u_{i,k} - u_{j,k}|}$. In other words, the more similar the two user attributes, the more likely that they are indistinguishable after transformation, and vice versa.

Definition 3. *(ϵ-**profile indistinguishability**) Given a set of user-attribute vectors $U = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$, a mechanism $\mathcal{M}(\cdot)$ satisfies ϵ-profile indistinguishability if and only if*

$$\frac{\Pr[\mathcal{M}(\boldsymbol{u}_i) = \widetilde{\boldsymbol{u}}]}{\Pr[\mathcal{M}(\boldsymbol{u}_j) = \widetilde{\boldsymbol{u}}]} \leq e^{\epsilon d(\boldsymbol{u}_i, \boldsymbol{u}_j)}, \tag{2}$$

for any pair of $\boldsymbol{u}_i, \boldsymbol{u}_j \in U$ and any possible perturbed user attribute vector $\widetilde{\boldsymbol{u}}$, where

$$d(\boldsymbol{u}_i, \boldsymbol{u}_j) = \sum_{k=1}^{\gamma} |u_{i,k} - u_{j,k}|$$

is the Manhattan distance between $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$.

The notion of ϵ-profile indistinguishability indicates that the ratio between the probabilities of two attribute vectors $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ each being transformed by mechanism $\mathcal{M}$ into the same $\widetilde{\boldsymbol{u}}$ is upper bounded by $e^{\epsilon d(\boldsymbol{u}_i, \boldsymbol{u}_j)}$. The more similar the two user attribute vectors, the smaller their Manhattan distance, the more likely that they are indistinguishable after transformation, and vice versa. We also note that both ϵ-attribute indistinguishability and ϵ-profile indistinguishability are weaker notions than ϵ-local differential privacy, as they only provide strong privacy guarantee for users with similar attributes and profiles, respectively.

B. Multivariate Laplace Mechanism

We now introduce a novel Multivariate Laplace Mechanism (MLM) which is based on two key observations. First, different user attributes have diverse sensitivity and thus desire different privacy protections. For example, home address is more sensitive than age or the time of last activity and thus requires higher level of privacy protection, i.e, a smaller privacy budget and larger amount of noise. Second, we find that adding independent noise to different attributes would result in perturbed user-attribute vectors widely dispersed, which would lead to lower classification accuracy and thus lower user-linkage accuracy. So our MLM adds positively correlated Laplace noise to different attributes. By carefully selecting the parameters of the Multivariate Laplace Distribution, our MLM achieves ϵ-attribute indistinguishability for individual attributes while providing high user-linkage accuracy.

We first briefly introduce the Multivariate Laplace Distribution [30]. Let $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_\gamma)$ be a γ-dimensional vector. The probability density function of a Multivariate Laplace Distribution $\mathcal{MLD}(\boldsymbol{\mu}, \Sigma)$ is given by

$$f(\boldsymbol{\psi}) = \frac{2}{(2\pi)^{\gamma/2}|\Sigma|^{1/2}} \cdot \frac{\mathcal{K}_{\gamma/2-1}(\sqrt{2q(\boldsymbol{\psi})})}{(\sqrt{q(\boldsymbol{\psi})/2})^{\gamma/2-1}}, \tag{3}$$

where $\mathcal{K}_{1-\gamma/2}(\cdot)$ is the modified Bessel function of the second kind with order $1-\gamma/2$ and $q(\boldsymbol{\psi}) = (\boldsymbol{\psi}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{\psi}-\boldsymbol{\mu})$. Substituting the Bessel function $\mathcal{K}_{1-\gamma/2}(\cdot)$ by the well-known asymptotic formula [30]

$$\mathcal{K}_{\gamma/2-1}(z) \approx \sqrt{\frac{\pi}{2z}}e^{-z}$$

if $|z| \to \infty$. We can derive the probability density function $f(\boldsymbol{\psi})$ as

$$f(\boldsymbol{\psi}) = \frac{2}{(2\pi)^{\gamma/2}|\Sigma|^{1/2}} \cdot \frac{(\frac{\pi}{2\sqrt{2\boldsymbol{\psi}^T\Sigma^{-1}\boldsymbol{\psi}}})^{1/2}\exp\left(-\sqrt{2\boldsymbol{\psi}^T\Sigma^{-1}\boldsymbol{\psi}}\right)}{(\sqrt{\frac{\boldsymbol{\psi}^T\Sigma^{-1}\boldsymbol{\psi}}{2}})^{\gamma/2-1}}, \tag{4}$$

where Σ is a $\gamma \times \gamma$ positive-definite covariance matrix given by

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{1,2} & \cdots & \sigma_1\sigma_\gamma\rho_{1,\gamma} \\ \sigma_2\sigma_1\rho_{2,1} & \sigma_2^2 & \cdots & \sigma_2\sigma_\gamma\rho_{2,\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\gamma\sigma_1\rho_{\gamma,1} & \cdots & \cdots & \sigma_\gamma^2 \end{bmatrix}.$$

The parameter σ_k in the above Σ is ψ_k's variance for all $1 \le k \le \gamma$, and $\rho_{i,j}$ is the correlation coefficient between ψ_i and ψ_j for all $1 \le i, j \le \gamma$. When $\rho_{i,j} \in (0,1)$ (or $(-1,0)$), ψ_i and ψ_j are positively (or negatively) correlated; when $\rho = 0$, they are uncorrelated.

We now detail the design of MLM. First, the OSN sets the privacy budget ϵ_k for the kth attribute according to its sensitivity for all $1 \le k \le \gamma$. Second, we compute the variance of each ψ_k as

$$\sigma_k = \frac{\sqrt{2}}{\epsilon_k}$$

Algorithm 1: Multivariate Laplace Mechanism(MLM)

Input: User attribute vector $\boldsymbol{u}_i$ and privacy budget $\epsilon_1, \ldots, \epsilon_\gamma$;
Output: Perturbed user attribute vector $\widetilde{\boldsymbol{u}}_i$;
foreach $k \in \{1, \ldots, \gamma\}$ **do**
  $\sigma_k \leftarrow \frac{\sqrt{2}}{\epsilon_k}$;
end
Compute a $\gamma \times \gamma$ covariance matrix Σ as in Eq. (5);
Draw a γ-dimensional vector $\boldsymbol{\psi}$ from $\mathcal{MLD}(\mathbf{0}, \Sigma)$;
$\widetilde{\boldsymbol{u}}_i \leftarrow \boldsymbol{u}_i + \boldsymbol{\psi}$;
return $\widetilde{\boldsymbol{u}}_i$;

for all $1 \le k \le \gamma$. Third, we set the covariance matrix as

$$\Sigma = \begin{bmatrix} \frac{2}{\epsilon_1^2} & \frac{2\rho_{1,2}}{\epsilon_1\epsilon_2} & \cdots & \frac{2\rho_{1,\gamma}}{\epsilon_1\epsilon_\gamma} \\ \frac{2\rho_{2,1}}{\epsilon_2\epsilon_1} & \frac{2}{\epsilon_2^2} & \cdots & \frac{2\rho_{2,\gamma}}{\epsilon_2\epsilon_\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{2\rho_{\gamma,1}}{\epsilon_\gamma\epsilon_1} & \frac{2\rho_{\gamma,2}}{\epsilon_\gamma\epsilon_2} & \cdots & \frac{2}{\epsilon_\gamma^2} \end{bmatrix} \tag{5}$$

where $\{\rho_{i,j} | 1 \le i, j \le \gamma\}$ are system parameters of which the impact is evaluated in Section V. Finally, given each user-attribute vector $\boldsymbol{u}_i \in U$, we draw a noise vector $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_\gamma)$ from the Multivariate Laplace Distribution $\mathcal{MLD}(\mathbf{0}, \Sigma)$ and output a perturbed vector as

$$\widetilde{\boldsymbol{u}}_i = \boldsymbol{u}_i + \boldsymbol{\psi}\ .$$

We summarize MLM in Algorithm 1.

C. Proposed Framework

Fig. 1 shows an overview of the proposed framework, which consists of three phases: data perturbation, model training, and user-linkage prediction.

In the data-perturbation phase, each participating OSN chooses its own privacy budgets $\epsilon_1, \ldots, \epsilon_\gamma$ according to the privacy sensitivity of each attribute. Consider OSN U as an example. It then perturbs its user-attribute vector set U using MLM to obtain a perturbed vector set $\widetilde{U}$. At the same time, U also seeks agreements from a subset of its users who are willing to share their data. This can be done through users' voluntary participating or offering certain rewards like online credits in exchange for their waiver of privacy protection [31]. Denote this subset of user-attribute vectors as $U^* \subset U$, where $|U^*| \ll |U|$. Then U sends both $\{(\boldsymbol{u}_i, \widetilde{\boldsymbol{u}}_i) | \boldsymbol{u}_i \in U^*\}$ and the remaining perturbed vectors $\{\widetilde{\boldsymbol{u}}_i | \boldsymbol{u}_i \in U \setminus U^*\}$ to the SDC. Similarly, OSN V submits $\{(\boldsymbol{v}_j, \widetilde{\boldsymbol{v}}_j) | \boldsymbol{v}_j \in V^*\}$ and the remaining perturbed vectors $\{\widetilde{\boldsymbol{v}}_j | \boldsymbol{v}_j \in V \setminus V^*\}$ to the SDC.

In the training phase, the SDC first constructs a training dataset from received user datasets. Since U^* and V^* both contain only original user-attribute vectors, the SDC can identify the users who appear in both U^* and V^* using standard methods, such as comparing the similarities of their social profiles, social activities, or other attributes. It is also possible for the SDC to adopt third party services such as

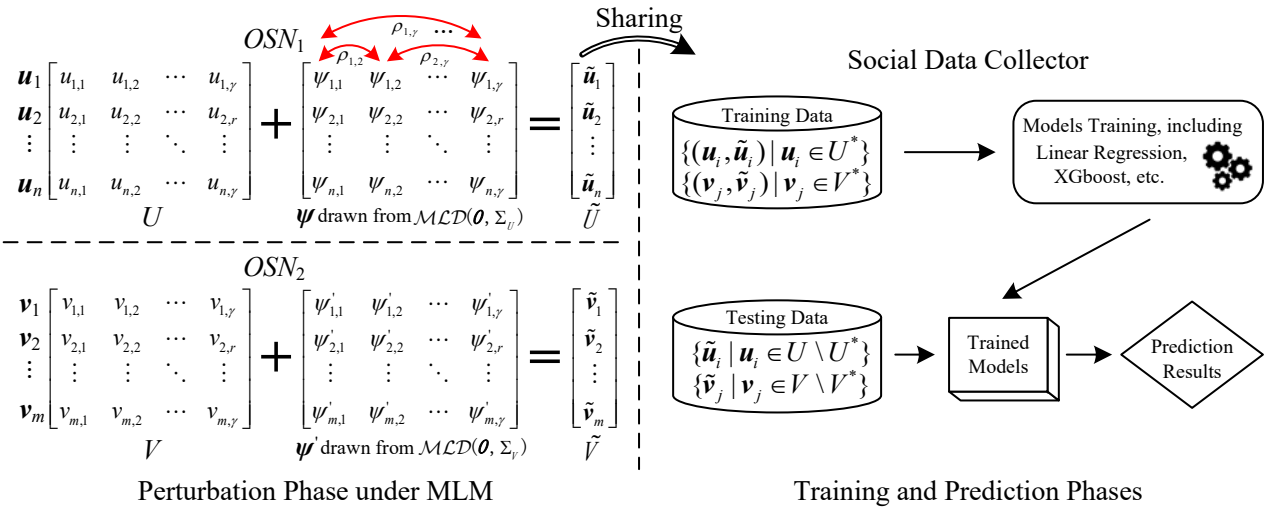


Fig. 1: Framework of differential privacy-preserving user linkage.

Everypost[3], Buffer[4], and Hootsuite[5] that allow users to link and synchronize multiple social media accounts. For every pair of user attribute vectors $(\boldsymbol{u}_i, \boldsymbol{v}_j)$ where $\boldsymbol{u}_i \in U$ and $\boldsymbol{v}_j \in V$, if they belong to the same real person, then the corresponding perturbed attribute vector pair $(\widetilde{\boldsymbol{u}}_i, \widetilde{\boldsymbol{v}}_j)$ is considered a positive instance. Otherwise, they are considered a negative instance. Subsequently, the SDC trains a classifier from the training dataset using standard machine learning techniques. In this paper, we consider and compare four machine learning methods, including *Linear Regression*, *XGboost*, *Adboost* with the linear kernel function, and *Logistic Regression*.

In the user-linkage prediction phase, for each pair of perturbed user attribute vectors $(\widetilde{\boldsymbol{u}}_i, \widetilde{\boldsymbol{v}}_j)$ such that $\boldsymbol{u}_i \in U \backslash U^*$ and $\boldsymbol{v}_j \in V \setminus V^*$, the SDC uses the trained classifier to determine whether they belong to the same real person.

V. Differential Privacy Analysis

In this section, we analyze the differential privacy guarantees offered by MLM. We first have the following theorem regarding the privacy protection for individual attributes.

Theorem 2. *The MLM mechanism $\mathcal{M}(\cdot)$ satisfies ϵ_k-attribute indistinguishability for all $1 \le k \le \gamma$.*

Proof. Consider any two attribute vectors $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,\gamma})$ and $\boldsymbol{u}_j = (u_{j,1}, \ldots, u_{j,\gamma})$. We abuse the notation to let $\mathcal{M}(u_{i,k})$ denote the kth attribute of perturbed vector output by $\mathcal{M}(\cdot)$ for all $1 \le k \le \gamma$. For any possible perturbed attribute $t_k \in \mathbb{R}$, we have

$$\begin{aligned}\frac{\Pr[\mathcal{M}(u_{i,k}) = t_k]}{\Pr[\mathcal{M}(u_{j,k}) = t_k]} &= \frac{\Pr[u_{i,k} + \psi_k = t_k]}{\Pr[u_{j,k} + \psi_k = t_k]} \\ &= \frac{\Pr[\psi_k = t_k - u_{i,k}]}{\Pr[\psi_k = t_k - u_{j,k}]}.\end{aligned} \tag{6}$$

[3]http://everypost.me/
[4]https://buffer.com/
[5]https://hootsuite.com/

Since the marginal distribution of each ψ_k of $\mathcal{MLD}(\mathbf{0}, \Sigma)$ is an Laplace distribution as $f(x) = \frac{1}{\sigma} e^{-\frac{|x|}{\sigma}}$ [30], and we set $\sigma = 1/\epsilon_k$, it follows that

$$\begin{aligned}\frac{\Pr[\psi_k = t_k - u_{i,k}]}{\Pr[\psi_k = t_k - u_{j,k}]} &= \frac{e^{\epsilon_k |t_k - u_{i,k}|}}{e^{\epsilon_k |t_k - u_{j,k}|}} \\ &= e^{\epsilon_k (|t_k - u_{i,k}| - |t_k - u_{j,k}|)} \\ &\le e^{\epsilon_k |u_{j,k} - u_{i,k}|}.\end{aligned} \tag{7}$$

□

We further have the following theorem regarding ϵ-profile indistinguishability of MLM.

Theorem 3. *MLM satisfies $\max(\{\epsilon_k\}_{k\in[1,\gamma]})$-profile indistinguishability if $\rho_{i,j} = 0$ for all $1 \le i, j \le \gamma$.*

Proof. Consider any two attribute vectors $\boldsymbol{u}_i = (u_{i,1}, \ldots, u_{i,\gamma})$ and $\boldsymbol{u}_j = (u_{j,1}, \ldots, u_{j,\gamma})$. Let $\boldsymbol{t} = \{t_1, \ldots, t_\gamma\}$ be an arbitrary perturbed attribute vector. If $\rho_{i,j} = 0$ for all $1 \le i, j \le \gamma$, then $\psi_1, \ldots, \psi_\gamma$ output by the MLM are independent of each other. We therefore have

$$\begin{aligned}\frac{\Pr[\mathcal{M}(\boldsymbol{u}_i) = \boldsymbol{t}]}{\Pr[\mathcal{M}(\boldsymbol{u}_j) = \boldsymbol{t}]} &= \frac{\prod_{k=1}^{\gamma} \Pr[\mathcal{M}_k(u_{i,k}) = t_k]}{\prod_{k=1}^{\gamma} \Pr[\mathcal{M}_k(u_{j,k}) = t_k]} \\ &= \frac{\prod_{k=1}^{\gamma} \Pr[u_{i,k} + \psi_k = t_k]}{\prod_{k=1}^{\gamma} \Pr[u_{j,k} + \psi_k = t_k]} \\ &= \frac{\prod_{k=1}^{\gamma} \Pr[\psi_k = t_k - u_{i,k}]}{\prod_{k=1}^{\gamma} \Pr[\psi_k = t_k - u_{j,k}]} \\ &= \frac{\prod_{k=1}^{\gamma} e^{\epsilon_k |t_k - u_{i,k}|}}{\prod_{k=1}^{\gamma} e^{\epsilon_k |t_k - u_{j,k}|}} \\ &= \prod_{k=1}^{\gamma} e^{\epsilon_k (|t_k - u_{i,k}| - |t_k - u_{j,k}|)} \\ &\le \prod_{k=1}^{\gamma} e^{\epsilon_k (|u_{j,k} - u_{i,k}|)} \\ &= e^{\sum_{k=1}^{\gamma} \epsilon_k (|u_{j,k} - u_{i,k}|)} \\ &\le e^{\max(\{\epsilon_k\}_{k\in[1,\gamma]}) d(\boldsymbol{u}_i, \boldsymbol{u}_j)}.\end{aligned} \tag{8}$$

□

We can further generalize the above theorem into user-attribute vector pairs across two OSNs. Assume that two OSN operators each choose $\{\epsilon_{1,1},\ldots,\epsilon_{1,\gamma}\}$ and $\{\epsilon_{2,1},\ldots,\epsilon_{2,\gamma}\}$ as their respective privacy budgets for γ attributes. The following theorem shows that any user attribute vector pair should be indistinguishable from another pair after perturbation.

Theorem 4. *Assume that two OSN operators each choose $\{\epsilon_{1,1},\ldots,\epsilon_{1,\gamma}\}$ and $\{\epsilon_{2,1},\ldots,\epsilon_{2,\gamma}\}$ as their respective privacy budgets for γ attributes. MLM satisfies ϵ-profile indistinguishability for profile pairs, where $\epsilon = \max(\{\epsilon_{j,k}\}_{j\in\{1,2\},1\le k\le\gamma})$.*

Proof. Consider any two pairs of user attribute vectors $(\boldsymbol{u}_{i_1},\boldsymbol{v}_{j_1})$ and $(\boldsymbol{u}_{i_2},\boldsymbol{v}_{j_2})$. Let $(\boldsymbol{u}_x,\boldsymbol{v}_y)$ be an arbitrary pair of perturbed user attribute vectors. We have

$$\begin{aligned}
&\frac{\Pr[\mathcal{M}(\boldsymbol{u}_{i_1})=\boldsymbol{u}_x,\mathcal{M}(\boldsymbol{v}_{j_1})=\boldsymbol{v}_y]}{\Pr[\mathcal{M}(\boldsymbol{u}_{i_2})=\boldsymbol{u}_x,\mathcal{M}(\boldsymbol{v}_{j_2})=\boldsymbol{v}_y]}\\
&=\frac{\Pr[\mathcal{M}(\boldsymbol{u}_{i_1})=\boldsymbol{u}_x]\cdot\Pr[\mathcal{M}(\boldsymbol{v}_{j_1})=\boldsymbol{v}_y]}{\Pr[\mathcal{M}(\boldsymbol{u}_{i_2})=\boldsymbol{u}_x]\cdot\Pr[\mathcal{M}(\boldsymbol{v}_{j_2})=\boldsymbol{v}_y]}\\
&\le e^{\max(\{\epsilon_{1,k}\}_{k\in[1,\gamma]})d(\boldsymbol{u}_{i_1},\boldsymbol{u}_{i_2})}\cdot e^{\max(\{\epsilon_{2,k}\}_{k\in[1,\gamma]})d(\boldsymbol{v}_{j_1},\boldsymbol{v}_{j_2})}\\
&\le e^{\max(\{\epsilon_{1,k},\epsilon_{2,k}\})(d(\boldsymbol{u}_{i_1},\boldsymbol{u}_{i_2})+d(\boldsymbol{v}_{j_1},\boldsymbol{v}_{j_2}))}\\
&= e^{\max(\{\epsilon_{1,k},\epsilon_{2,k}\})d((\boldsymbol{u}_{i_1},\boldsymbol{v}_{j_1}),(\boldsymbol{u}_{i_2},\boldsymbol{v}_{j_2}))}.
\end{aligned} \tag{9}$$

□

Furthermore, we examine the relationship between ϵ-profile indistinguishability and ϵ-text indistinguishability [8], which is a similar privacy notion defined over the Euclidean distance. ϵ-text indistinguishability is defined over text vectors, which are transformed into real-value vectors as mentioned in Section II-A. In particular, a mechanism $\mathcal{M}(\cdot)$ satisfies ϵ-text indistinguishability if and only if

$$\frac{\Pr[\mathcal{M}(\boldsymbol{u}_i)=\widetilde{\boldsymbol{u}}]}{\Pr[\mathcal{M}(\boldsymbol{u}_j)=\widetilde{\boldsymbol{u}}]}\le e^{\epsilon\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)},$$

where $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$ are any two user-attribute vectors, and $\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)$ is their Euclidian distance. The following theorem establishes the connection between these two notions.

Theorem 5. *If a mechanism $\mathcal{M}$ satisfies ϵ-profile indistinguishability, then it also satisfies $\sqrt{\gamma}\epsilon$-text indistinguishability.*

Proof. Consider any two γ-dimensional user attribute vectors $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. Let $d(\boldsymbol{u}_i,\boldsymbol{u}_j)$ and $\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)$ be their Manhattan and Euclidian distances, respectively. By definition, we have

$$d(\boldsymbol{u}_i,\boldsymbol{u}_j)=\sum_{k=1}^{\gamma}|u_{i,k}-u_{j,k}|$$

and

$$\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)=(\sum_{k=1}^{\gamma}(u_{i,k}-u_{j,k})^2)^{\frac{1}{2}}.$$

Let $\triangle u_k=|u_{i,k}-u_{j,k}|$ for all $1\le k\le\gamma$. It follows that

$$\frac{d^2(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\mathsf{d}^2(\boldsymbol{u}_i,\boldsymbol{u}_j)}=\frac{(\sum_{k=1}^{\gamma}\triangle u_k)^2}{\sum_{k=1}^{\gamma}\triangle u_k^2}.$$

The above ratio takes the maximum value γ when $\triangle u_1=\cdots=\triangle u_\gamma$. It follows that the maximum ratio $\frac{d(\boldsymbol{u}_i,\boldsymbol{u}_j)}{\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)}$ is $\sqrt{\gamma}$, and therefore $d(\boldsymbol{u}_i,\boldsymbol{u}_j)\le\sqrt{\gamma}\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)$. Now assume that a mechanism $\mathcal{M}$ satisfies ϵ-profile indistinguishability. We have

$$\frac{\Pr[\mathcal{M}(\boldsymbol{u}_i)=\boldsymbol{t}]}{\Pr[\mathcal{M}(\boldsymbol{u}_j)=\boldsymbol{t}]}\le e^{\epsilon d(\boldsymbol{u}_i,\boldsymbol{u}_j)}\le e^{\epsilon\sqrt{\gamma}\mathsf{d}(\boldsymbol{u}_i,\boldsymbol{u}_j)}. \tag{10}$$

□

We also analyze the impact of injected noise on user-attribute vectors. Theorem 6 estimates the expected Manhattan distance between a user attribute vector and the corresponding perturbed one under MLM.

Theorem 6. *Let $\boldsymbol{u}_i$ and $\widetilde{\boldsymbol{u}}_i$ be an original user attribute vector and the corresponding perturbed output by MLM, the expected Manhattan distance between $\boldsymbol{u}_i$ and $\widetilde{\boldsymbol{u}}_i$ is $\sum_{k=1}^{\gamma}\frac{2}{\epsilon_k}$, where ϵ_k is the privacy budget of the k-th attribute.*

Proof. Let $\boldsymbol{\psi}=(\psi_1,\ldots,\psi_\gamma)$ be the noise vector output by the MLM. The Manhattan distance between $\boldsymbol{u}$ and $\widetilde{\boldsymbol{u}}$ is given by $d(\boldsymbol{u}_i,\widetilde{\boldsymbol{u}}_i)=\sum_{k=1}^{\gamma}|\psi_k|$ Since $\boldsymbol{\psi}$ is sampled from $\mathcal{MLD}(\mathbf{0},\Sigma)$, and the marginal probability distribution of each ψ_k is $f(\psi_k)=\frac{1}{\sigma_k}e^{-\frac{|\psi_k|}{\sigma_k}}$, we have

$$\begin{aligned}
\mathbb{E}(|\psi_k|)&=\int_{-\infty}^{\infty}\epsilon_k e^{-\epsilon_k|\psi_k|}|\psi_k|d\psi_k\\
&=2\epsilon_k\int_0^{\infty}e^{-\epsilon_k\psi_k}\psi_k d\psi_k\\
&=2\epsilon_k(-\frac{1}{\epsilon_k}\psi_k e^{-\epsilon_k\psi_k}|_0^{\infty}-\frac{1}{\epsilon_k^2}e^{-\epsilon_k\psi_k}|_0^{\infty})\\
&=\frac{2}{\epsilon_k}.
\end{aligned}$$

It follows that

$$\mathbb{E}(d(\boldsymbol{u}_i,\widetilde{\boldsymbol{u}}_i))=\mathbb{E}(\sum_{k=1}^{\gamma}|\psi_k|)=\sum_{k=1}^{\gamma}\mathbb{E}(|\psi_k|)=\sum_{k=1}^{\gamma}\frac{2}{\epsilon_k}.$$

□

VI. Experiment Results

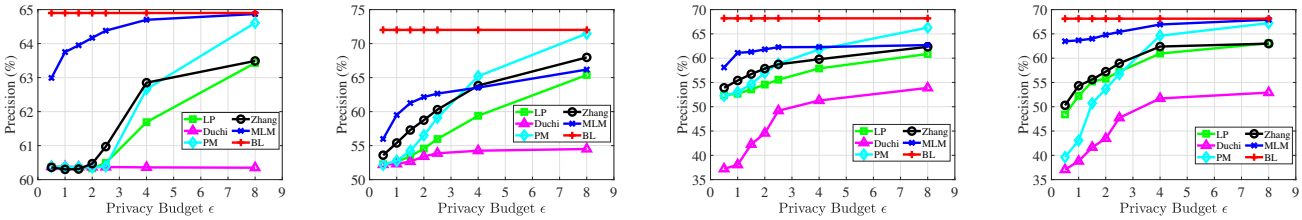
A. Datasets, Parameter Setting, and Performance Metrics

We use three public OSN datasets published in [32], including `Twitter`, `MySpace` and `Last.fm`. The number of users in the `Twitter`, `MySpace` and `Last.fm` datasets are 28,199, 9,993, and 7,661, respectively. Each user has 6, 9, and 9 attributes in three social datasets, respectively. We also use the 19,126 and 5,002 pairs of users as the ground-truth for two linked social datasets `Twitter-MySpace` and `MySpace-Last.fm`, which were originally collected by Perito *et al*. [33] through the Google Profiles service.

We preprocess the datasets by converting all text data into numeric data. Specifically, we first utilize the standard 2-gram

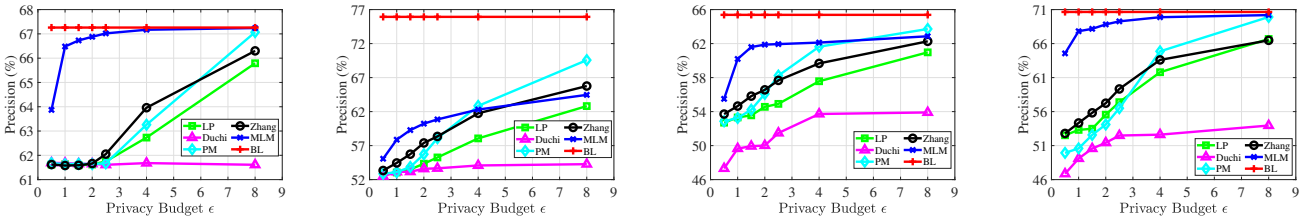
TABLE I: User-Linkage Prediction Performance.

	Twitter-MySpace				MySpace-Last.fm			
	Precision	AUC	Recall	F_1	Precision	AUC	Recall	F_1
Linear Rregression	0.6494	0.5689	0.5668	0.5736	0.6742	0.6003	0.5848	0.6087
XGboost	0.6669	0.5445	0.6436	0.5691	0.6231	0.5841	0.6412	0.5988
Adboost	0.7061	0.5650	0.6504	0.5680	0.6898	0.5835	0.6699	0.6092
Logistic Rregression	0.6861	0.5672	0.6496	0.5742	0.7100	0.5848	0.6742	0.6087



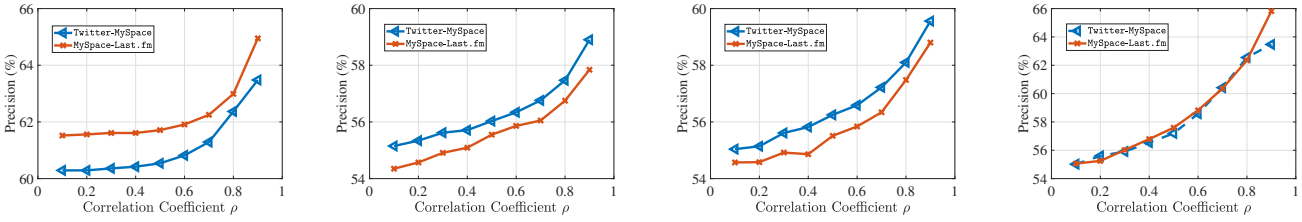
(a) *Linear Regression* (b) *XGBoost* (c) *AdBoost* (d) *Logistic Regression*

Fig. 2: Prediction accuracy on the Twitter-MySpace dataset.



(a) *Linear Regression* (b) *XGBoost* (c) *AdBoost* (d) *Logistic Regression*

Fig. 3: Prediction accuracy on the MySpace-Last.fm dataset.



(a) *Linear Regression* (b) *XGBoost* (c) *AdBoost* (d) *Logistic Regression*

Fig. 4: Impact of ρ on different machine learning methods.

to convert each text string into a set of substrings and then use the Simhash library [25] to compute a hash value for the substring set. As we mentioned in Section II, SimHash is locality sensitive, which guarantees that two similar text strings must have similar SimHash values, and vice versa.

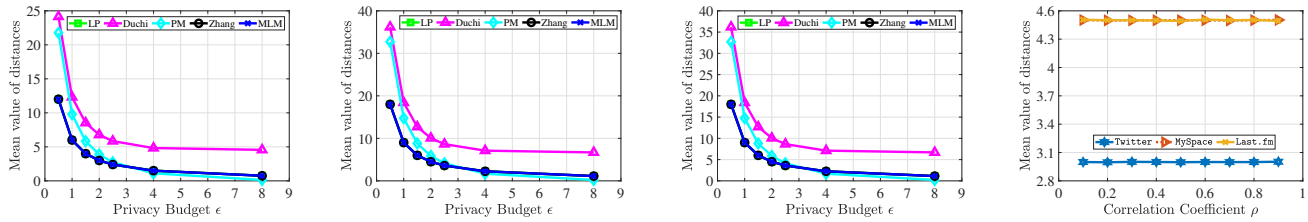
We use the following parameters in our experiments. We set the privacy budget $\epsilon_k = \epsilon$ for all $1 \leq k \leq \gamma$, where ϵ is chosen from $\{0.5, 1, 1.5, 2, 2.5, 4, 8\}$. We also set $\rho_{i,j} = \rho$ for all $1 \leq i, j \leq \gamma$, where ρ is chosen from $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$.

Four machine learning methods are evaluated in our framework, including *Linear Regression*, *XGboost*, *Adboost* with the linear kernel function, and *Logistic Regression*. We also compare our framework with Laplace Mechanism [34], Duchi [35], Piecewise Mechanism [36], and Zhang's work [8]. The performance metrics include Precision, Area under the ROC Curve (AUC), Recall, and the F_1 score.

B. Performance of Prediction without User Privacy

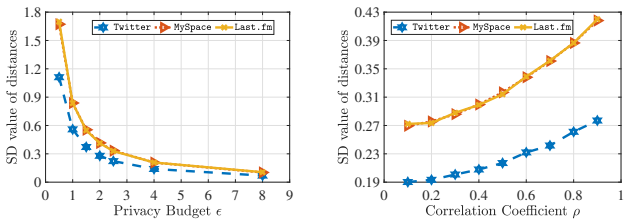
Table I lists user-linkage prediction results on two original dataset pairs without considering user privacy. As we can see, the prediction accuracy of the four machine learning methods on different datasets is different. This is reasonable because the same machine learning technique typically exhibits different performance on different datasets.

Due to similar results and space limitation, we do not present all the performance metrics but only show the precision under various machine learning methods. Figs. 2 and 3 show the precision with different perturbation mechanisms, where each point represents the average of 100 runs, each with a random perturbation, and the correlation coefficient ρ under MLM is set to 0.9. We can see that the precision increases as



(a) Dataset `Twitter` (b) Dataset `MySpace` (c) Dataset `Last.fm` (d) Different ρ

Fig. 5: Mean Manhattan distance vs. privacy budget ϵ.



(a) Different ϵ (b) Different ρ

Fig. 6: Standard deviation of Manhattan distance vs. privacy budget ϵ and correlation coefficient ρ.

ϵ raises. This is expected because the larger the privacy budget ϵ, the smaller the injected noise, the higher precision, and vice versa. Besides, the precision of *Linear Regression* and *Logistic Regression* methods under MLM is significantly higher than that under other four mechanisms, including Laplace Mechanism [34], Duchi [35], Piecewise Mechanism [36], and Zhang's work [8]. Moreover, the precision of *XGBoost* and *AdBoost* under MLM is only slightly higher than that under other four mechanisms. This is expected, as the correlation of injected noise under MLM is larger than that under other mechanisms, and *Linear Regression* and *Logistic Regression* methods are more sensitive to the correlation among injected noise than *XGBoost* and *AdBoost*. To evaluate the impact of ρ on the precision under MLM, we fix ϵ to 2 and vary the correlation coefficient ρ from 0.1 to 0.9. Fig. 4 shows that the precision increases as ρ increases from 0.1 to 0.9. In addition, when ρ is close to 0, the precision is the lowest. Generally speaking, the stronger the correlation of injected noise, the better the prediction accuracy, and vice versa. In summary, the privacy budget ϵ and the correlation coefficient ρ both affect the prediction performance.

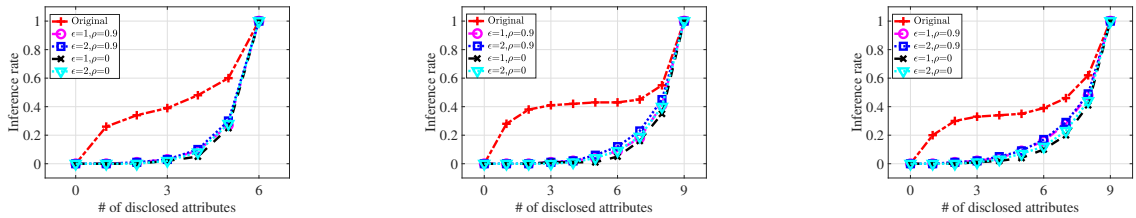
Furthermore, we evaluate the impact of ρ and ϵ on the mean of Manhattan distance between perturbed and original user-attribute vectors. First, we fix ρ as 0.9 and vary ϵ from 0.5 to 8. As we can see from Figs. 5a, 5b, and 5c, the mean Manhattan distance between perturbed and original attribute vectors under our MLM is smaller than that under other perturbation mechanisms when ϵ is smaller than 4. In addition, the mean Manhattan distance under MLM is larger than that under PM [36], which is also similar to what we have seen in Figs. 2 and 3. This is also anticipated, as the smaller the injected noise, the smaller the Manhattan distance between the original and perturbed user-attribute vectors, and vice versa. Second, we fix ϵ as 2 and vary ρ from 0.1 to 0.9 to evaluate the impact of ρ on the mean Manhattan distance. Fig. 5d shows that the mean Manhattan distance between perturbed and original user-attribute vectors is not affected by the change in ρ, which coincides with Theorem 5. Moreover, the mean distance for `MySpace` and `Last.fm` datasets are the same and higher than that in `Twitter`. This is reasonable because the more attributes, the smaller the privacy budget, the larger the mean Manhattan distance, and vice versa.

Finally, we evaluate the impact of ρ and ϵ on the standard deviation of the Manhattan distance between perturbed and original user-attribute vectors. We first fix ρ to 0.9 and vary ϵ from 0.5 to 8 and then fix ϵ to 2 and vary ρ from 0.1 to 0.9. Figs. 6a and 6b show that the standard deviation of the Manhattan distance decreases as ϵ increases and increases as ρ increases. Moreover, the results for `MySpace` and `Last.fm` datasets are the same and higher than that for the `Twitter` dataset. The reason is that that both `MySpace` and `Last.fm` datasets have 9 user attributes while the `Twitter` dataset has only 6 user attributes.

C. Defense Against User-Linkage Attack

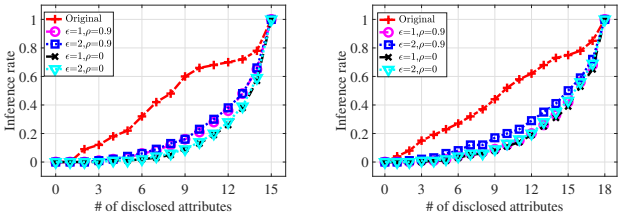
Our mechanism is developed for defending against the user-linkage attack. The notions of ϵ-attribute indistinguishability and ϵ-profile indistinguishability in Definitions 2 and 3 indicate that similar attributes (or profiles) should be perturbed into the same attribute (or profile) with similar probabilities. By satisfying the two notions, our mechanism makes it harder for the attacker to carry out the user-linkage attack.

We conduct the following experiment to evaluate the effectiveness of our mechanism against the user-linkage attack. For each user $\boldsymbol{u}$ in one dataset, we first calculate its K-nearest neighbor set denoted by $\mathcal{S}_{\boldsymbol{u}}$. We then model the strength of the attacker by assuming that he/she knows κ randomly chosen attribute values of user $\boldsymbol{u}$, where κ ranges from 0 to γ for an individual victim and from 0 to $\sum_{k=1}^{2}\gamma_k$ for a pair of linked users. Next, we calculate the K-nearest neighbor set of $\boldsymbol{u}$ with respect to the κ known attributes, denoted by $\mathcal{S}_{\boldsymbol{u}'}$. Finally, we calculate the inference rate as $|\mathcal{S}_{\boldsymbol{u}} \bigcap \mathcal{S}_{\boldsymbol{u}'}|/K$. An inference of one indicates that the attacker is able to identify the user from the κ known attributes. We repeat the above process 100 times for each user in the dataset each with a different subset of κ attributes and calculate the average rate over all the users.



(a) Dataset `Twitter` (b) Dataset `MySpace` (c) Dataset `Last.fm`

Fig. 7: Performance of user-linkage attack on original datasets.



(a) `Twitter-MySpace` (b) `MySpace-Last.fm`

Fig. 8: Performance of user-linkage attack on linked dataset.

Figs. 7a to 7c show the inference rate for `Twitter`, `MySpace` and `Last.fm` datasets with $\langle \epsilon, \rho \rangle$ set to $\langle 1, 0 \rangle$, $\langle 1, 0.9 \rangle$, $\langle 2, 0 \rangle$, and $\langle 2, 0.9 \rangle$, where κ ranges from 0 to 6 and $K = 2$. We can see that the inference rate increases as κ increases in all five cases. For example, the inference rate under the perturbed `MySpace` dataset increases from 12% to 45% as κ increases from 6 to 8. This is expected, as the more attributes the adversary knows, the more likely he/she can identify the real user. Moreover, the inference rate under the proposed mechanism is much lower than that for the original dataset without perturbation. In addition, the inference rate increases as ϵ increases for the same ρ. The reason is that the smaller the ϵ, the more likely two similar attributes being perturbed to the same one, the stronger defense against the user-linkage attack, and vice versa. Similarly, the inference rate increases as ρ increases for the same ϵ as the larger the ρ, the stronger correlations of among attributes, the weaker defense against the user-linkage attack. In addition, Fig. 8 shows the inference rate for two linked datasets with $K = 2$. We can draw similar conclusions as those from `Twitter`, `MySpace` and `Last.fm` datasets. We omit the detailed discussions here due to the space limitations.

VII. Related Work

A. Local Differential Privacy

LDP [13] is a strong and rigorous mathematical privacy-preserving framework that provides semantic and information-theoretic guarantees on individuals' privacy. Compared to the centralized differential privacy model, LDP assumes that the data curator is semi-trusted or even malicious.

Early work on LDP focuses on statistical estimation. Erlingsson *et al.* [27] introduced a Randomized Aggregatable Privacy-Preserving Ordinal Response (RAPPOR) mechanism for collecting statistics over a set of binary values with LDP guarantee, which requires the dictionary to be known in advance. To overcome this limitation, Fanti *et al.* [37] proposed an improved mechanism to support more sophisticated statistics like joint distribution. Duchi *et al.* [14] studied two kinds of statistical estimators under LDP, including mean estimation and convex risk minimization. Kairouz *et al.* [38] introduced the staircase mechanism to maximize data utility, which can be reduced to solving a linear optimization problem. Gu *et al.* [39] devoted to the correlation of key-value pairs, and proposed a novel framework PCKV with two protocols (i.e., PCKV-UE and PCKVGRR) whereby to tighten privacy budget and achieve better utility.

Recent years have also witnessed the success of LDP applied in various data analytics problems, including probability distribution estimation [14], heavy hitter discovery [15], percentile statistics [16], frequent new term discovery [17], frequency estimation [18], frequent itemset mining [19], marginal release [20], clustering [21], location privacy [22], [23], and so on. Since we tackle a totally different problem, none of these works are directly applicable.

Directly applying LDP to multi-dimension OSN data suffers from the curse of dimensionality [8]. To overcome these limitations, Andrés *et al.* [29] first proposed the notion of ϵ-geo-indistinguishability, which provides weaker privacy protection than LDP by taking the distance between two locations into account. Later, Zhang *et al.* [8] introduced a related privacy notion ϵ-text indistinguishability for privacy-preserving social data publishing. Both of these works satisfy the corresponding privacy notions by adding a noise vector with length drawn from a Laplace distribution, but they cannot satisfy indistinguishability for individual attributes.

B. Privacy-Preserving Data Publishing

Privacy-preserving data publishing has been extensively studied. Classical techniques include k-anonymity [10], ℓ-diversity [11] and t-closeness [12]. Therein, k-anonymity is proposed to defend against re-identification attacks, in which every user identity should be indistinguishable with at least $k - 1$ other users' identities. However, it cannot prevent probabilistic attack [40] in which an attacker infers sensitive information without recovering the user's identity. ℓ-diversity and t-closeness were proposed to address the limitations of k-anonymity. These privacy-preserving mechanisms can achieve

satisfactory privacy protection for well-defined relational data but lack a rigorous theoretical framework.

VIII. Conclusion

In this paper, we studied the problem of differential privacy-preserving user linkage across multiple OSNs. In view of the limitation of LDP, we first introduce two novel privacy notions of ϵ-attribute indistinguishability and ϵ-profile indistinguishability, which offer privacy protection while taking the distance between user-attribute vectors into account. We further presented a novel Multivariate Laplace Mechanism (MLM) to achieve these privacy notions and a differential privacy-preserving user linkage framework based on the MLM and machine learning techniques. Detailed theoretical analysis confirmed the privacy guarantee of the proposed MLM, and experimental studies using three real datasets demonstrated the significant advantages of our framework over prior solutions.

Acknowledgment

This work was partially supported by National Natural Science Foundation of China through grants 61902433, Hunan Provincial Natural Science Foundation of China through grants 2019JJ50802 and 2019JJ50288, US National Science Foundation through grants CNS-1933069, CNS-1824355, CNS-1651954 (CAREER), CNS-1718078 and CNS-1933047.

References

[1] A. Malhotra, L. Totti, W. Meira, P. Kumaraguru, and V. Almeida, "Studying user footprints in different online social networks," in *ASONAM'12*, Istanbul, Turkey, Aug. 2012.

[2] M. Yan, J. Sang, C. Xu, and M. Hossain, "A unified video recommendation by cross-network user modeling," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 4, p. 53, Aug. 2016.

[3] J. Zhang, P. Yu, and Z. Zhou, "Meta-path based multi-network collective link prediction," in *KDD'14*, New York, NY, Aug. 2014.

[4] O. Goga, P. Loiseau, R. Sommer, R. Teixeira, and K. Gummadi, "On the reliability of profile matching across large online social networks," in *KDD'15*, Sydney, Australia, Aug. 2015.

[5] A. Andreou, O. Goga, and P. Loiseau, "Identity vs. Attribute Disclosure Risks for Users with Multiple Social Profiles," in *ASONAM'17*, Sydney, Australia, July 2017.

[6] K. Buraya, A. Farseev, A. Filchenkov, and T. Chua, "Towards user personality profiling from multiple social networks," in *AAAI'17*, San Francisco, CA, Feb. 2017.

[7] T. Chen, M. Kaafar, A. Friedman, and R. Boreli, "Is more always merrier?: a deep dive into online social footprints," in *WOSN'12*, Helsinki, Finland, Aug. 2012.

[8] J. Zhang, J. Sun, R. Zhang, Y. Zhang, and X. Hu, "Privacy-preserving social media data outsourcing," in *INFOCOM'18*, Honolulu, HI, Oct. 2018.

[9] X. Meng, S. Wang, K. Shu, J. Li, B. Chen, H. Liu, and Y. Zhang, "Personalized privacy-preserving social recommendation," in *AAAI'18*, New Orleans, LA, Feb. 2018.

[10] L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557 – 570, Oct. 2002.

[11] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "ℓ-diversity: Privacy beyond k-anonymity," in *ICDE'06*, Atlanta, GA, Apr. 2006.

[12] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy Beyond k-Anonymity and ℓ-Diversity," in *ICDE'07*, Istanbul, Turkey, Apr. 2007.

[13] S. Kasiviswanathan, H. Lee, K. Nissim, S. Raskhodnikova, and A. Smith, "What can we learn privately?" *SIAM Journal on Computing*, vol. 40, no. 3, pp. 793 – 826, June 2011.

[14] J. Duchi, M. Jordan, and M. Wainwright, "Local privacy and statistical minimax rates," in *FOCS'13*, Berkeley, CA, Oct. 2013.

[15] R. Bassily, K. Nissim, U. Stemmer, and A. Thakurta, "Practical locally private heavy hitters," in *NIPS'17*, Long Beach, CA, Dec. 2017.

[16] J. Sun, R. Zhang, J. Zhang, and Y. Zhang, "PriStream: Privacy-preserving distributed stream monitoring of thresholded percentile statistics," in *INFOCOM'16*, San Francisco, CA, Apr. 2016.

[17] N. Wang, X. Xiao, Y. Yang, T. Hoang, H. Shin, J. Shin, and G. Yu, "PrivTrie: Effective frequent term discovery under local differential privacy," in *ICDE'18*, Paris, France, Apr. 2018.

[18] R. Bassily and A. Smith, "Local, private, efficient protocols for succinct histograms," in *STOC'15*, Portland, OR, June 2015.

[19] T. Wang, N. Li, and S. Jha, "Locally differentially private frequent itemset mining," in *S&P'18*, San Francisco, CA, May. 2018.

[20] G. Cormode, T. Kulkarni, and D. Srivastava, "Marginal release under local differential privacy," in *SIGMOD'18*, Houston, TX, June 2018.

[21] K. Nissim and U. Stemmer, "Clustering algorithms for the centralized and local models," in *ALT'18*, Lanzarote, Spain, Apr. 2018.

[22] X. Jin, R. Zhang, Y. Chen, T. Li, and Y. Zhang, "DPSense: Differentially private crowdsourced spectrum sensing," in *CCS'16*, Vienna, Austria, Oct. 2016.

[23] X. Jin and Y. Zhang, "Privacy-preserving crowdsourced spectrum sensing," *IEEE/ACM Transactions on Networking*, vol. 26, no. 3, pp. 1236 – 1249, June 2018.

[24] W. Cavnar and J. Trenkle, "N-gram-based text categorization," in *SDAIR'94*, Las Vegas, NV, Apr. 1994.

[25] C. Sadowski and G. Levin, "Simhash: Hash-based similarity detection," *Technical report, Google*, 2007.

[26] Y. Wang, C. Si, and X. Wu, "Regression model fitting under differential privacy and model inversion attack," in *IJCAI'15*, Buenos Aires, Argentina, July 2015.

[27] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *CCS'14*, Scottsdale, AZ, Nov. 2014.

[28] S. Su, P. Tang, X. Cheng, R. Chen, and Z. Wu, "Differentially private multi-party high-dimensional data publishing," in *ICDE'16*, Helsinki, Finland, May 2016.

[29] M. Andrés, N. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-Indistinguishability: Differential Privacy for Location-Based Systems," in *CCS'13*, Berlin, Germany, Nov. 2013.

[30] S. Kotz, T. Kozubowski, and K. Podgorski, *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, Dec. 2012.

[31] S. Palan and C. Schitter, "Prolific. ac—A subject pool for online experiments," *Journal of Behavioral and Experimental Finance*, vol. 17, pp. 22 – 27, 2018.

[32] Y. Zhang, J. Tang, Z. Yang, J. Pei, and P. Yu, "COSNET: Connecting Heterogeneous Social Networks with Local and Global Consistency," in *KDD'15*, Sydney, Australia, Aug. 2015.

[33] D. Perito, C. Castelluccia, M. Kaafar, and P. Manils, "How unique and traceable are usernames?" in *PETS'11*, Waterloo, Canada, July 2011.

[34] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *TCC'06*, New York, NY, Mar. 2006.

[35] J. Duchi, M. Jordan, and M. Wainwright, "Minimax optimal procedures for locally private estimation," *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 182 – 201, 2018.

[36] N. Wang, X. Xiao, Y. Yang, J. Zhao, S. Hui, H. Shin, J. Shin, and G. Yu, "Collecting and Analyzing Multidimensional Data with Local Differential Privacy," in *ICDE'19*, Macau Sar, China, Apr. 2019.

[37] G. Fanti, V. Pihur, and Ú. Erlingsson, "Building a rappor with the unknown: Privacy-preserving learning of associations and data dictionaries," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 3, pp. 41 – 61, July 2016.

[38] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," in *NIPS'14*, Montreal, Canada, Dec. 2014.

[39] X. Gu, M. Li, Y. Cheng, L. Xiong, and Y. Cao, "PCKV: Locally Differentially Private Correlated Key-Value Data Collection with Optimized Utility," in *USENIX Security'20*, Virtual Conference, Aug. 2020.

[40] K. Wang, R. Chen, B. Fung, and P. Yu, "Privacy-preserving data publishing: A survey on recent developments," *ACM Computing Surveys*, vol. 42, no. 4, pp. 141 – 153, 2010.