Capturing the Effects of Attribute based Correlation on Privacy in Micro-databases

Debanjan Sadhya[1], Bodhi Chakraborty[2] and Sanjay Kumar Singh[1]
[1]*Dept. of Computer Science and Engineering, Indian Institute of Technology (Banaras Hindu University), Varanasi, India*
[2]*Dept. of Information Technology, Indian Institute of Information Technology Allahabad, Allahabad, India*

Keywords: Privacy, Linking Attack, Micro-database, Information Theory.

Abstract: In the modern data driven era, it is a very common practice for individuals to provide their personalized data in multiple databases. However, the existence of correlated information in between these databases is a common source of privacy risk for the database users. In our study, we investigate such scenarios for attribute based linking attacks. These attacks refer to the common strategy by which an adversary can breach the privacy of the database respondents via exploiting the correlated information among the database attributes. In our work, we have proposed an information theoretic framework through which the achievable privacy levels following an adversarial linking attack are quantified. Our developed model also incorporates various aspects associated with micro-databases such as sanitization mechanism and auxiliary side information, thereby providing a more holistic structure to our theoretical framework. A comparative analysis of the various cases associated with our model theoretically confirms the notion that a sanitization mechanism facilitates in preserving the original privacy levels of the users.

1 INTRODUCTION

Datasets which contain person specific information about individual respondents are termed as micro-databases. Based upon the nature of the data which they represent, the database attributes can be categorized into three categories namely Identifiers, Key attributes (or quasi identifiers) and confidential (sensitive) attributes (Rebollo-Monedero et al., 2010). Identifiers are those attributes which unambiguously identify the respondents. Typical examples of these in micro-databases include 'SSN number' and 'passport number'. These values are either removed or encrypted prior to distribution due to the high privacy risks associated with them. Key attributes are those properties which can be linked or combined with external sources or databases to re-identify a respondent. Typical examples of such attributes include 'age', 'gender' and 'address'. Sensitive attributes contain the most critical data of the users; maintaining their confidentiality is the primary objective of any database security scheme. Examples for these type of attributes include 'medical diagnosis', 'political affiliation' and 'salary'.

Accumulating person specific data in some central storage facility poses great risks for the individuals participating in the data collection process. Although there are well defined policies and guidelines to restrict the types of publishable data (Fung et al., 2010), they are regularly circumvented in practical data sharing scenarios. As a direct consequence, the likelihood for an adversary to easily retrieve some critical information about any targeted individual from the databases remains alarmingly high. From the perspective of the adversary, the most effective method for obtaining critical information about a target is via exploiting the correlation among the public attributes (quasi identifiers) in multiple databases. Alternatively, it can be said that there exist multiple 'links' among databases which subsequently assists an adversary in performing various malicious attacks. These type of attacks are commonly known as linking, cross-matching or correlation based attacks. The severity of the linking attacks comprehensively increases if the adversary possesses some auxiliary background information about the targeted individual.

In this paper, we have attempted to formally capture the effects of linking attacks on the privacy levels of the database respondents. Intuitively speaking, the privacy of an individual decreases in the event that an adversary can successfully establish attribute based linkages. There are two major contri-

Sadhya, D., Chakraborty, B. and Singh, S.
Capturing the Effects of Attribute based Correlation on Privacy in Micro-databases.
DOI: 10.5220/0006417504300436
In *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - Volume 4: SECRYPT*, pages 430-436
ISBN: 978-989-758-259-2

butions of our work. Firstly, we have provided theoretical frameworks for modeling micro-databases, along-with attribute based links, sanitization mechanisms and adversarial background knowledge. Subsequently, we have quantified the amount of resulting privacy following a successful linking attack based on a couple of micro-databases. Our study is motivated by the pioneering work of (Sankar et al., 2013), in which the authors gave an information theoretical analysis of the trade-off between utility and privacy in micro databases. The rest of this paper is organized as follows. Section 2 introduces some background concepts which subsequently form the basis of our analytic work, and Section 3 contains the detailed construction of our formal framework for micro-databases. The process of quantifying the privacy after a linking attack is described in Section 4, and finally Section 5 concludes our work with potential future scopes.

2 BACKGROUND CONCEPTS

In this section we briefly discuss about multiple aspects related to privacy and micro-databases. These inter-related background notions would facilitate in the development process of our formal model.

2.1 Privacy

Traditionally, privacy has been used as a metric for measuring the level of uncertainty of information corresponding to an individual within a database. Privacy preservation was first described by (Dalenius, 1977) as the guarantee that an adversary learns nothing extra about any target if the adversary gains access to published data. Regarding databases, public and private attributes are generally modeled as random variables having a specific joint probability distribution. The privacy of an individual remains intact (i.e. there is no privacy loss) if the disclosure of the associated public attributes provide no additional information about the corresponding private attributes. In a probabilistic sense, it can be stated that the conditional entropy of the private attribute should remain as high as possible after an adversary observes the public attributes. Conventionally, privacy has been accounted for in an information theoretic way. The uncertainty about a piece of undisclosed information is related to its information content. The information content of a source S is measured by its entropy H which is defined as-

$$H(S) = \sum_i p_i \log \frac{1}{p_i}$$

where p_i is the probability with which a character s_i is emitted from the source S.

Let X_{prv} denote the set of random variables which represent sensitive attributes in a database. Similarly, let X_{PUB} characterize the set of random variables corresponding to the public information which is accessible by the adversary. As demonstrated in subsequent sections, the source of this information can be external public attributes, as well as correlated auxiliary side information. Furthermore, let's assume that X_{prv} and X_{PUB} are correlated by a joint probability distribution function $p_{(X_{prv},X_{PUB})}(y,x)$ where $\forall(y,x)|y \in X_{prv}, x \in X_{PUB}$. Under such a naive scenario, the privacy $(\mathcal{P})$ can be quantified as -

$$\mathcal{P} = H(X_{prv}|X_{PUB})$$

where $H(X_{prv}|X_{PUB})$ represents the conditional entropy (equivocation) of X_{prv} given X_{PUB}. The parameter $\mathcal{P}$ accurately captures the essence of privacy as it represents the leftover entropy of the private data on the disclosure of the associated public data. An equivalent metric privacy risk (R) (Rebollo-Monedero et al., 2010) is defined as the mutual information between the public and private random variables. Thus,

$$R = I(X_{prv};X_{pub}) = H(X_{prv}) - H(X_{prv}|X_{pub})$$

Both privacy risk (R) and privacy $(\mathcal{P})$ are complementary and essentially capture the same notion. However in our work we would define privacy by the equivocation $\mathcal{P}$.

2.2 Linking Attacks

Privacy breaches through linking can occur by three separate mechanisms namely *record linkage*, *attribute linkage* and *table linkage*. These three techniques correspond to the situation when an attacker is able to link an individual to a record in a published data table, to a sensitive attribute in a published data table, or to the published data table itself respectively. Regarding record and attribute linkages, it is assumed that the attacker has the prior knowledge that the targeted individual's record is present in the database. However the objective of the attacker in table linkage based attacks is to determine whether the individual's record is present or absent in the released table.

Our present work is concerned only with *attribute linkages* since the most famous and influential attacks in the real world were carried out by linking correlated attributes. Prominent examples of attribute based linking attacks include re-identifying

the sensitive medical record of William Weld (governor of Massachusetts) by joining it with public voter databases (Sweeney, 2005),(Sweeney, 1997); de-anonymization of individual DNA sequences (Malin and Sweeney, 2004) and privacy breaches owing to AOL search data (Hansell, 2006). Perhaps the most influential work was done by (Narayanan and Shmatikov, 2006) where de-anonymization of Netflix subscribers was performed by correlating it with external data obtained from Internal Movie Database (IMDB). This study resulted in identification of Netflix records of known users, thereby revealing their apparent political preferences and other related sensitive information.

2.3 Sanitization Mechanisms

Almost all the micro-data datasets are published in the public domain after removing the *Identifiers* associated with the database subjects. This process is known as *anonymization*. However even after anonymization, an adversary can query critical information about the subjects by the virtue of their public attributes present in the dataset. This problem motivated for the development of techniques which suppress the disclosure risk of individual information as much as possible while maximizing the utility of the published data. These privacy preserving mechanisms are generally termed as sanitization processes. Broadly speaking, there are three main approaches for performing sanitization prior to publishing the micro-data datasets. Although the methods for achieving privacy are different, their underlying concept is the same - modification of the original data that is to be released. These techniques are termed as *generalization*, *anatomization* and *perturbation* (Fung et al., 2010).

Generalization techniques alter the original data so that they cannot be identified later on. These methods calculate a common value for a group of records and then replace the individual records contained within the groups with the computed common value. Prominent privacy preservation schemes which employ generalization as their underlying notion include k-anonymity (Samarati and Sweeney, 1998), t-closeness (Machanavajjhala et al., 2007) and l-diversity (Li et al., 2007). Anatomization algorithms (Xiao and Tao, 2006) disassociate the relationship between quasi identifiers and sensitive attributes. This method partitions the original data into two separate tables - one containing only quasi identifiers and the other one consisting solely of sensitive attributes. However both the tables contain a connecting attribute termed as *GroupID*. All records in an equivalent group will have the same value of *GroupID* in both tables, and therefore remain linked to the sensitive values in the group in the exact same way.

Offering an alternate solution for achieving data privacy is the perturbation method. The driving principle behind this technique is the addition of external noise to the original data to produce a synthetic output. However the perturbation must be carried out specifically such that any statistical information computed from the original database must not significantly differ from the same statistical information estimated form the synthetic output. Although exhaustively researched, perhaps the most famous example of perturbation based sanitization mechanism is Differential privacy (Dwork, 2006). This method is generally based on a query-response framework wherein external noise sampled from a pre-defined distribution is added to the statistics of the original database.

As already discussed, micro-databases are released in the public domain only after sanitizing them through one of the discussed mechanisms. However, the records of the database can also be obtained (by the adversary) in their original form in the event of a database leakage. Hence to make our model more practical, we have taken into consideration all the distinct possibilities regarding the sanitization procedure while quantifying the attainable privacy levels in Section 4.

2.4 Auxiliary Background Knowledge

In addition to the public information available, an adversary can also utilize any background information related to the subject. Incorporating this aspect gives more power to the adversary for mining sensitive information about individuals. A famous example of this notion can be described by the commonly referred Terry Gross's height (Dwork, 2006). It basically states that supposing 'height' was considered a sensitive information, an adversary possessing the background knowledge that "Terry Gross is two inches shorter than the average Lithuanian women", can accurately calculate Terry Gross's height from a statistical database containing average heights of woman of different nationalities. Inclusion of background information is crucial for the construction of any real-world data dependent framework since it accurately captures the adversarial model. Moreover it provides a realistic estimate of privacy limits since it has been showed that absolute privacy protection is not possible due to presence of related background information (Dalenius, 1977). Intuitively it can be understood that the privacy of an individual decreases with the amount of background information

possessed by the adversary. This observation stems from the fact that privacy is inversely proportional to the net amount of disclosed information about an individual.

3 MODEL CONSTRUCTION

This section is dedicated towards formally modeling a generic micro-database along-with its related dependencies. We have developed our models considering the availability of two micro-databases since the majority of real-world attacks on privacy were carried out involving a couple of databases. For instance, the Netflix de-anonymization (Narayanan and Shmatikov, 2006) was executed utilizing Netflix and Internet Movie Database (IMDb).

3.1 Assumptions

Prior to initiating our formal constructions, we first present some assumptions which we have made regarding the distribution and correlation of attributes in a micro-database. These suppositions assist us in developing a formal and consistent mathematical model. All these assumptions have been already justified and subsequently used in previous works (Sankar et al., 2013). Firstly, we model a micro-database as a collection of n observations (rows) generated by a memoryless source whose outputs are independently and identically distributed (i.i.d). Additionally, the rows of the database is a collection of correlated attributes that is generated according to its probability of occurrence from a well defined source. Some assumptions are also made with respect to the adversary. We consider him/her to possess some auxiliary background information regarding either any particular targeted individual or the entire group of subjects in the database. For instance, the adversary may know whether or not an individual had participated in a database. This assumption enables us not only to take into consideration the various possibilities of privacy breach, but also makes our model more generic.

3.2 Micro-database Model

We start by defining the notations for two micro-databases DB^1 and DB^2. [1] Let K^1 and K^2 denote the number of attributes in the two databases; also let $\mathcal{K}^1$ and $\mathcal{K}^2$ be the sets representing these attributes. Let $X^1_{\mathcal{K}}$ and $X^2_{\mathcal{K}}$ denote the set of random variables representing the attributes of the two databases respectively, thus $X^1_{\mathcal{K}} = \{X^1_i : i = 1,2,...K^1\}$ and $X^2_{\mathcal{K}} = \{X^2_i : i = 1,2,...K^2\}$. Let DB^1 and DB^2 consist of n independent observations (i.e. rows) which follow joint probability distributions -

$$p_{X^1_{\mathcal{K}}}(x_{\mathcal{K}^1}) = p_{X^1_1 X^1_2 ... X^1_{K^1}}(x_1, x_2, ...x_{K^1})$$

and

$$p_{X^2_{\mathcal{K}}}(x_{\mathcal{K}^2}) = p_{X^2_1 X^2_2 ... X^2_{K^2}}(x_1, x_2, ...x_{K^2})$$

The above dependencies captures the correlation between the attributes in the corresponding databases. However in accordance to previous works, we assume independence among the rows.

Let $\mathcal{K}^1_{pub}; \mathcal{K}^1_{prv}$ and $\mathcal{K}^2_{pub}; \mathcal{K}^2_{prv}$ represent public and private attributes in the two databases respectively. It should be noted that $(\mathcal{K}^1_{pub} \cup \mathcal{K}^1_{prv}) = \mathcal{K}^1$,$(\mathcal{K}^1_{pub} \cap \mathcal{K}^1_{prv}) = \phi$, $(\mathcal{K}^2_{pub} \cup \mathcal{K}^2_{prv}) = \mathcal{K}^2$ and $(\mathcal{K}^2_{pub} \cap \mathcal{K}^2_{prv}) = \phi$. We further denote the set of their corresponding random variables by $X^1_{\mathcal{K}_{pub}}, X^1_{\mathcal{K}_{prv}}, X^2_{\mathcal{K}_{pub}}$ and $X^2_{\mathcal{K}_{prv}}$ respectively. Thus,

$$X^1_{\mathcal{K}_{pub}} = \{X^1_i\}_{i \in \mathcal{K}^1_{pub}}; X^1_{\mathcal{K}_{prv}} = \{X^1_i\}_{i \in \mathcal{K}^1_{prv}}$$

$$X^2_{\mathcal{K}_{pub}} = \{X^2_i\}_{i \in \mathcal{K}^2_{pub}}; X^2_{\mathcal{K}_{prv}} = \{X^2_i\}_{i \in \mathcal{K}^2_{prv}}$$

3.3 Attribute based Correlation

Now we proceed in expressing attribute based correlation between the two databases DB^1 and DB^2. For introducing similarity, we assume that some of the public attributes from both the databases overlap. This assumption is practical since real world micro-databases normally contains interrelated attributes. We restrict the type of overlapping attributes to public (and not private) since linking attacks are based solely on public attributes. Let the number of these common attributes be denoted by K°. Since both the databases are distinct, $K^\circ < min(K^1, K^2)$. Let these attributes be represented as a set $\mathcal{K}^\circ$, thus $|\mathcal{K}^\circ| = K^\circ$ and $\mathcal{K}^\circ = \{\mathcal{K}^1_{pub} \cap \mathcal{K}^2_{pub}\}$. Accordingly, let the random variable representing $\mathcal{K}^\circ$ be denoted by $X_{\mathcal{K}^\circ}$. Let these common records follow the joint probability distribution -

$$p_{X_{\mathcal{K}^\circ}}(x_{\mathcal{K}^\circ}) = p_{X^\circ_1 X^\circ_2 ... X^\circ_{K^\circ}}(x_1, x_2, ...x_{K^\circ})$$

Thus this distribution essentially captures the attribute based association between the two databases.

[1] We will refer to properties of the first and second databases with superscripts [1] and [2] respectively

3.4 Sanitization Process

We present generic sanitization mechanisms on both the databases. Accordingly, we define encoding functions F_1 and F_2 which maps DB^1 and DB^2 to a set of indices $J^1 = \{1,2,...M^1\}$, $J^2 = \{1,2,...M^2\}$ and a set of associated output sanitized databases SDB^1 and SDB^2. Here M^1 and M^2 denotes the number of sanitized databases for DB^1 and DB^2 respectively. Thus,

$$F_1 : DB^1 \rightarrow J^1, \{SDB_k^1\}_{k=1}^{M^1}$$

and

$$F_2 : DB^2 \rightarrow J^2, \{SDB_k^2\}_{k=1}^{M^2}$$

This encoding function is a little different to that previously used (Sankar et al., 2013), in the sense that our functions maps only the databases, thereby making the encoding a one-to many function. Additionally we do not require the decoding function since we are not concerned about the *utility* of the databases.

3.5 Background Knowledge

For our framework, the background information is modeled as n-length sequences and denoted by the random variables Z^1 and Z^2 corresponding to DB^1 and DB^2 respectively. Thus,

$$Z^1 = (Z_1^1, Z_2^1, ...Z_n^1) \quad \text{and} \quad Z^2 = (Z_1^2, Z_2^2, ...Z_n^2)$$

where (Z_i^1, Z_i^2) take values from a finite set $\mathcal{Z}$.

Also, let the side information corresponding to the correlated attributes be represented by Z°. It should be noted that it is not necessary that $Z^\circ \subseteq (Z^1 \cup Z^2)$, i.e. the correlation among the attributes might reveal some additional background information to the adversary. On the other hand, the side information itself must be correlated with the databases to be meaningful. These correlations are denoted by the joint probability distribution functions $p_{X^1_{\mathcal{K}} Z^1}(x_{\mathcal{K}^1}, z^1)$ and $p_{X^2_{\mathcal{K}} Z^2}(x_{\mathcal{K}^2}, z^2)$ corresponding to DB^1 and DB^2 respectively.

4 PRIVACY LEVELS AND LINKING ATTACKS

In this section, we formally quantify the privacy guarantees following a successful execution of attribute based linking attacks by an adversary. As mentioned previously, privacy is defined as the reduction in entropy of sensitive information given that an adversary has access to some correlated public information. Since in our work we deal with two databases, privacy quantification is done in two levels. In the first level we estimate the privacy loss on account of private attributes present in both databases, whereas in the second level we approximate the further reduction in privacy while considering the correlated attributes for the databases. This provides a hierarchical mechanism for calculating the final privacy loss corresponding to the micro-database subjects. Based on the implementation of a sanitization procedure, we can formulate three distinct cases regarding the mechanism of privacy loss. These cases correspond to the scenarios when - (i) None of the databases are sanitized, (ii) Only one of the databases is sanitized, and (iii) Both the databases are sanitized. The final privacy level in each case is denoted by $\mathcal{P}^i$, where i denotes the case number.

4.1 No Database is Sanitized

First we consider only DB^1. Since the private attributes are available to the adversary in their original form, privacy ($\mathcal{P}_1$) is given by -

$$\mathcal{P}_1 = H(X^1_{\mathcal{K}_{prv}} | X^1_{\mathcal{K}_{prv}}, Z^1) \geq E_1$$

where $\mathcal{P}_1$ is lower bounded by E_1. For this case, the value of $\mathcal{P}_1$ equates to 0. This observation is consistent with the intuition that the micro-database subjects would have no privacy in case their private attribute values are accessible by the adversary (via a leakage). Moreover, the side information correlated with the first database would have no significance in this first level of privacy quantification as the privacy cannot be further reduced from 0. Next we consider the second database DB^2. Similar to the previous case, the privacy ($\mathcal{P}_2$) is quantified as -

$$\mathcal{P}_2 = H(X^2_{\mathcal{K}_{prv}} | X^2_{\mathcal{K}_{prv}}, Z^2) \geq E_2$$

where E_2 is a general lower bound on $\mathcal{P}_2$. The quantity $\mathcal{P}_2$ also equates to 0 since the private attributes of the second database are also available to the adversary in the unaltered form.

In level two, we determine the effects of correlated attributes on the level 1 privacy states. Essentially, we estimate the remaining entropy of the two databases when the adversary performs linking attacks on the basis of the overlapped attributes. For such a case, the quantities $\mathcal{P}_1$ and $\mathcal{P}_2$ serve as the maximum amount of remaining information (privacy) for the database subjects. Moreover since $\mathcal{P}_1$ is a function of $(X^1_{\mathcal{K}_{prv}}, Z^1)$, it can be represented by a random variable $X_{\mathcal{P}_1}$ with the mapping-

$$X_{\mathcal{P}_1} : (X^1_{\mathcal{K}_{prv}}, Z^1) \rightarrow [E_1, H(X^1_{\mathcal{K}_{prv}})]$$

Similarly $\mathcal{P}_2$ can be represented by a random variable $X_{\mathcal{P}_2}$ with the mapping -

$$X_{\mathcal{P}_2} : (X^2_{\mathcal{K}_{prv}}, Z^2) \rightarrow [E_2, H(X^2_{\mathcal{K}_{prv}})]$$

In this special case, the random variables $X_{\mathcal{P}_1}$ and $X_{\mathcal{P}_2}$ are defined on the range $\{0\}$ since $\mathcal{P}_1, \mathcal{P}_2 = 0$. Let the leftover privacy involving DB^1 and DB^2 after observing the correlated attributes be denoted by $\mathcal{P}_3$ and $\mathcal{P}_4$ respectively. Thus -

$$\mathcal{P}_3 = H(X_{\mathcal{P}_1} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_3$$

and

$$\mathcal{P}_4 = H(X_{\mathcal{P}_2} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_4$$

where E_3 and E_4 are general lower bounds on $\mathcal{P}_3$ and $\mathcal{P}_4$ respectively. However in this particular case both $\mathcal{P}_3$ and $\mathcal{P}_4$ equate to 0, since the random variables $X_{\mathcal{P}_1}$ and $X_{\mathcal{P}_2}$ are defined on $\{0\}$.

Hence the total leftover privacy for this case equates to -

$$\mathcal{P}^1 = \mathcal{P}_3 + \mathcal{P}_4 = 0 \tag{1}$$

4.2 Only One Database is Sanitized

For the sake of simplicity, we assume that DB^1 is sanitized whereas DB^2 is not. The privacy quantification process for the alternative assumption is simply the symmetrically opposite case (i.e. the notations for DB^1 and DB^2 gets interchanged). Since the first database is sanitized in this case, the only attack strategy of the adversary is to obtain sensitive information from the sanitized database and the related side information. Thus the privacy of the subject ($\mathcal{P}_1$) equates to -

$$\mathcal{P}_1 = H(X^1_{\mathcal{K}_{prv}} | J^1, Z^1) \geq E_1$$

To reiterate, J^1 is the index of the sanitized database corresponding to DB^1. The maximum value of $\mathcal{P}_1$ occurs when (J^1, Z^1) reveals no information about $X^1_{\mathcal{K}_{prv}}$, i.e. when $X^1_{\mathcal{K}_{prv}}$ is independent of both J^1 and Z^1. Privacy in that case equates to the entropy of $X^1_{\mathcal{K}_{prv}}$,i.e. $H(X^1_{\mathcal{K}_{prv}})$. However the second database is available in the original format, and consequently the adversary is able to directly extract all critical information from there. In such a case the privacy ($\mathcal{P}_2$) becomes -

$$\mathcal{P}_2 = H(X^2_{\mathcal{K}_{prv}} | X^2_{\mathcal{K}_{prv}}, Z^2) \geq E_2$$

As in the previous case, $\mathcal{P}_2$ equates to 0. Now we begin the quantification process for level 2. First we represent $\mathcal{P}_1$,$\mathcal{P}_2$ as random variables $X_{\mathcal{P}_1}$,$X_{\mathcal{P}_2}$ with the following mapping functions -

$$X_{\mathcal{P}_1} : (X^1_{\mathcal{K}_{prv}}, J^1, Z^1) \rightarrow [E_1, H(X^1_{\mathcal{K}_{prv}})]$$

and

$$X_{\mathcal{P}_2} : (X^2_{\mathcal{K}_{prv}}, Z^2) \rightarrow [E_2, H(X^2_{\mathcal{K}_{prv}})]$$

Subsequently, the second level privacy equates to -

$$\mathcal{P}_3 = H(X_{\mathcal{P}_1} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_3$$

and

$$\mathcal{P}_4 = H(X_{\mathcal{P}_2} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_4$$

In this case, $\mathcal{P}_4$ equates to 0 since $X_{\mathcal{P}_2}$ is defined on $\{0\}$. Hence the final remaining privacy equates to -

$$\mathcal{P}^2 = \mathcal{P}_3 + \mathcal{P}_4 = H(X_{\mathcal{P}_1} | X_{\mathcal{K}^\circ}, Z^\circ) \tag{2}$$

4.3 Both Databases are Sanitized

This final case accounts for the majority of practical scenarios since micro-databases are generally sanitized prior to public distribution. In this case, the adversary is able to obtain sensitive information about the subjects on the basis of attribute based linking attacks. The amount of meaningful information which the adversary is able to mine from them depends upon the effectiveness of the sanitization mechanism. The level 1 privacy for DB^1 and DB^2 are denoted by -

$$\mathcal{P}_1 = H(X^1_{\mathcal{K}_{prv}} | J^1, Z^1) \geq E_1$$

and

$$\mathcal{P}_2 = H(X^2_{\mathcal{K}_{prv}} | J^2, Z^2) \geq E_2$$

Due to the effects of sanitization, both $\mathcal{P}_1, \mathcal{P}_2 \neq 0$. They are subsequently represented by the random variables $X_{\mathcal{P}_1}$ and $X_{\mathcal{P}_2}$, which are defined by the mappings -

$$X_{\mathcal{P}_1} : (X^1_{\mathcal{K}_{prv}}, J^1, Z^1) \rightarrow [E_1, H(X^1_{\mathcal{K}_{prv}})]$$

and

$$X_{\mathcal{P}_2} : (X^2_{\mathcal{K}_{prv}}, J^2, Z^2) \rightarrow [E_2, H(X^2_{\mathcal{K}_{prv}})]$$

Subsequently, the level 2 privacy for DB^1 and DB^2 are represented by -

$$\mathcal{P}_3 = H(X_{\mathcal{P}_1} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_3$$

and

$$\mathcal{P}_4 = H(X_{\mathcal{P}_2} | X_{\mathcal{K}^\circ}, Z^\circ) \geq E_4$$

Similar to the level 1 privacy (i.e. $\mathcal{P}_1$ and $\mathcal{P}_2$), both $\mathcal{P}_3, \mathcal{P}_4 \neq 0$. Thus the final privacy levels can be quantified as -

$$\mathcal{P}^3 = \mathcal{P}_3 + \mathcal{P}_4 = H(X_{\mathcal{P}_1}|X_{\mathcal{K}^\circ}, Z^\circ) + H(X_{\mathcal{P}_2}|X_{\mathcal{K}^\circ}, Z^\circ) \quad (3)$$

On comparing the values of $\mathcal{P}^1$, $\mathcal{P}^2$ and $\mathcal{P}^3$ from Eqn. 1, Eqn. 2 and Eqn. 3 respectively, we can represent a ordinal relationship among them as -

$$\mathcal{P}^1 < \mathcal{P}^2 < \mathcal{P}^3$$

This relationship follows from the facts that $\mathcal{P}^1 = 0$ and $H(X_{\mathcal{P}_2}|X_{\mathcal{K}^\circ}, Z^\circ)$ is a positive quantity. The maximum amount of privacy gets preserved when we implement appropriate sanitization procedures on both the databases, whereas the total privacy attains the lower bound of 0 (i.e. no privacy is preserved) when none of the databases are sanitized. Hence this relation also vindicates the notion that a sanitization mechanism facilitates in preserving the user's privacy.

5 CONCLUSION AND FUTURE SCOPES

In our work, we have attempted to formally quantify the achievable privacy levels in the lieu of attribute based linking attacks involving micro-databases. We have taken into consideration the various possibilities in which an adversary may try to learn sensitive information about an individual and provided the appropriate levels of privacy in each case. Additionally, we have computed the privacy levels for three distinct cases based on the application of a sanitization mechanism on the micro-database. Our findings theoretically confirm the intuitive notion that a sanitization procedure assists in preserving the privacy of the database respondents.

Privacy breaches in micro-databases primarily occur due to the existence of multiple attribute based links among the records of the databases. Although our work successfully models this setting, the only constraint of our work is related to the number of available micro-databases (to the adversary). More specifically speaking, we have assumed that an adversary is able to perform the linking based attacks on the basis of two micro-databases. Modifying our framework for incorporating more than two micro-databases is a natural extension of our work. Finally we would like to experimentally evaluate our framework on real-life datasets, which would provide empirical validation of our study.

REFERENCES

Dalenius, T. (1977). Towards a methodology for statistical disclosure control. *Statistik Tidskrift*, 15(429-444):2–1.

Dwork, C. (2006). Differential privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, pages 1–12, Berlin, Heidelberg. Springer-Verlag.

Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv.*, 42(4):14:1–14:53.

Hansell, S. (2006). Aol removes search data on vast group of web users. Technical report, New York Times.

Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1).

Malin, B. and Sweeney, L. (2004). How (not) to protect genomic data privacy in a distributed network: Using trail re-identification to evaluate and design anonymity protection systems. *J. of Biomedical Informatics*, 37(3):179–192.

Narayanan, A. and Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. *CoRR*, abs/cs/0610105.

Rebollo-Monedero, D., Forne, J., and Domingo-Ferrer, J. (2010). From t-closeness-like privacy to postrandomization via information theory. *IEEE Trans. on Knowl. and Data Eng.*, 22(11):1623–1636.

Samarati, P. and Sweeney, L. (1998). Generalizing data to provide anonymity when disclosing information (abstract). In *Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, PODS '98, pages 188–, New York, NY, USA. ACM.

Sankar, L., Rajagopalan, S. R., and Poor, H. V. (2013). Utility-privacy tradeoffs in databases: An information-theoretic approach. *IEEE Transactions on Information Forensics and Security*, 8(6):838–852.

Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine & Ethics*, 25(2-3):98–110.

Sweeney, L. (2005). Statement before the privacy and integrity advisory committee of the department of homeland security. Technical report, Department of Homeland Security.

Xiao, X. and Tao, Y. (2006). Anatomy: Simple and effective privacy preservation. In *Proceedings of the 32Nd International Conference on Very Large Data Bases*, VLDB '06, pages 139–150. VLDB Endowment.