Secure k-Anonymization Linked with Differential Identifiability (Workshop)

Zheng Zhao[1(✉)], Tao Shang[2], and Jianwei Liu[2]

[1] School of Electronic and Information Engineering, Beihang University, Beijing 100083, China
zhaozheng1000@163.com

[2] School of Cyber Science and Technology, Beihang University, Beijing 100083, China
shangtao@buaa.edu.cn

Abstract. Most k-anonymization mechanisms that have been developed presently are vulnerable to re-identification attacks, e.g., those generating a generalized value based on input databases. k-anonymization mechanisms do not properly capture the notion of hiding in a crowd, because they do not impose any constraints on the mechanisms. In this paper, we define (k, ρ)-anonymization that achieves secure k-anonymization notion linked with differential identifiability under the condition of privacy parameter ρ. Both differential identifiability and k-anonymization limit the probability that an individual is re-identified in a database after an adversary observes the output results of the database. Furthermore, differential identifiability can provide the same strong privacy guarantees as differential privacy. It can make k-anonymization perform securely, while (k, ρ)-anonymization achieves the relaxation of the notion of differential identifiability, which can avoid a lot of noise and help obtain better utility for certain tasks. We also prove the properties (k, ρ)-anonymization under composition that can be used for application in data publishing and data mining.

Keywords: Differential identifiability · k-anonymization · Privacy preservation

1 Introduction

Privacy-preserving notions for data publishing and data mining have achieved many advances with the increase of collected data that is used for various data analysis. Many privacy definitions and applications for releasing data securely have been introduced in the literatures (see [16] and [20] for surveys). k-anonymity was proposed by Sweeny and Samarati [17–19] to protect the content of released data records. Some follow-up notions include l-diversity [15] and t-closeness [13]. The most prominent is k-anonymity. Its basic idea is to ensure that each quasi-identifier group has at least k tuples in order that individuals cannot be uniquely re-identified. The notion of k-anonymity tries to work on the attributes of quasi-identifiers, which is exposed to some subtle but effective attacks. Even the k-anonymity that treats all attributes as quasi-identifiers

H. Gao et al. (Eds.): ChinaCom 2019, LNICST 313, pp. 307–316, 2020.
https://doi.org/10.1007/978-3-030-41117-6_25

does not provide sufficient privacy preservation against re-identification attacks [9]. In addition, k-anonymization mechanisms that satisfy k-anonymity have weaknesses, because they do not properly capture the notion of hiding in a crowd. Thus it is necessary to solve such problem and define a new notion of k-anonymization that is secure without related weaknesses.

A privacy notion that is widely accepted is differential privacy (DP) developed by a series of works [1–5]. The basic idea is that any individual in a database has only a limited influence on the output of the database to hide the contribution of any single individual. Since privacy is a social notion with many facets, research fields examine various facets of privacy to understand strength and weakness. In 2011, Kifer and Machanavajjhala [11] argued that differential privacy is not robust to arbitrary background knowledge and impossible to provide privacy and utility without the assumption of data. Gehrke et al. [8] introduced zero-knowledge to provide sufficient protection when an individual may be strongly correlated with other individuals. In 2012, Lee and Clifton [12] considered that differential privacy does not match legal definitions of privacy, which is required to protect individually identifiable data. As a result, they proposed differential identifiability (DI). The privacy parameter ρ limits the probability estimate that is the contribution of an individual to the output results.

Although privacy definitions have some differences, both differential identifiability and differential privacy provide strong privacy guarantees. In fact, such strong privacy definitions are not suitable for all scenarios. Some mechanisms may add a lot of noise to satisfy the given privacy definition, thus reducing the utility of released data. For differential privacy, some relaxation notions have been proposed. In 2012, Li et al. [14] proposed the definition of differential privacy under sampling that captures the adversary's uncertainty about the input database. The results showed that sampling is a powerful tool that can greatly benefit differential privacy when sampling is used correctly. Gehrke et al. [7] introduced crowd-blending privacy that can achieve better utility and strictly relax the notion of differential privacy. However, for differential identifiability, there does not exist any relaxation or even properties on it for application in data publishing and data mining. So the new notion of k-anonymization should be able to achieve the relaxation of the notion of differential identifiability.

In order to provide sufficient privacy guarantees, we further study k-anonymization privacy preservation mechanisms and its link with differential identifiability. Both differential identifiability and k-anonymization limit the probability that an individual is re-identified in a database after an adversary observes the output results of the database. In this paper, we define a new notion of (k, ρ)-anonymization that makes k-anonymization linked with differential identifiability to properly capture the notion of hiding in a crowd. The new notion can make k-anonymization perform securely and achieve the relaxation of the notion of differential identifiability. Furthermore, we prove the properties of (k, ρ)-anonymization under composition that can allow us to apply it for

complex privacy preservation issues in data mining and data publishing. Taken together, the results can provide basis for practical application of differential identifiability.

2 Preliminaries

A database D can be considered a finite multiset. Each attribute value is a fixed value in the universe U. Each entry in U can correspond to an individual in the database that privacy should be protected. $I(t)$ denotes the identity of the individual corresponding to the entry t in U. $\mathcal{I}_D = \{I(t)|t \in D\}$ denotes the set of individuals which belong to D. $D' \subset D$ is a database having one less individual than D, i.e., $|D'| = |D| - 1$.

Lee and Clifton [12] argued that differential privacy limits how much one individual can affect an output, not how much information can be leaked about an individual. This does not match legal definitions of privacy, which is required to protect individually identifiable data. Thus they proposed the definition of ρ-differential identifiability that can provide the same guarantees as differential privacy, but ρ limits the probability estimate that an individual belongs to the input database. The definition is:

Definition 1 *(ρ-differential identifiability [12]). A randomized mechanism M is said to satisfy ρ-differential identifiability if for all databases D, any $D' = D - t^*$, for any entry $t \in U - D'$:*

$$Pr[I(t) \in \mathcal{I}_D | M(D) = R, D'] \leq \rho. \tag{1}$$

The definition of ρ-differential identifiability limits the identifiability risk of any individual in the universe U, thus the posterior probability that any individual t belonging to the database is less than or equal to ρ after an adversary observes the output response R. In order to calculate the posterior probability, it is necessary to assume prior beliefs that an adversary may have.

To measure the adversary's confidence in making an inference, the proposed definition assumes that there exists a *possible worlds model* [12] in which the adversary considers the set of all possible databases. Given the adversary's prior knowledge $\mathcal{L} = \langle U, D', \mathcal{I}'_D \rangle$, the set of all possible databases Ψ is

$$\Psi = \{D' \cup \{t\} | t \in U \wedge t \notin D'\}.$$

Every possible world $\omega \in \Psi$ is equally likely to be D. Only one of the databases in Ψ is the true database which generates the output response R. In other words, only one individual is uncertain, and this individual must be drawn uniform from $m = |\Psi| = |U| - |D'|$ possible individuals with the probability between 0 and 1. At the same time, Lee and Clifton have experimentally proved that when the value of ρ is close to the correct probability of a random guess, the output response is barely utility and the privacy goal is also violated. Thus ρ-differential identifiability will be useful when $\rho > \frac{1}{m}$.

3 New Secure k-Anonymization Notion Linked with Differential Identifiability

Both differential identifiability and differential privacy provide strong privacy guarantees. For some data analysis tasks, some mechanisms may add a lot of noise to satisfy the given privacy definition while reducing the utility of released data. Furthermore, such strong privacy guarantees may be too restrictive for specific data analysis. We may require a privacy definition that can be strictly relaxed. In this section, we focus on a new k-anonymization notion linked with differential identifiability, which can relax the notion of differential identifiability and be more secure than classical k-anonymization.

3.1 Classical k-Anonymization

k-anonymity [19] is a privacy definition specifically for protecting data records of tables. A published table satisfies k-anonymity if each quasi-identifier (QID) group has at least k records in the table to reduce the probability of identification. k-anonymity requires the separation of all attributes into quasi-identifiers (QIDs) and sensitive attributes (SAs). The adversary is assumed to only know QIDs. Such separation is very hard to achieve in practice. And any separation between QIDs and SAs based on the adversary's background knowledge can be easily violated. There may exist an adversary that knows sensitive information on some individuals. If these individuals can be re-identified based on these information, it is still a privacy leak.

The literature [14] makes a clear distinction between k-anonymity and k-anonymization algorithms. k-anonymization aims to generate the anonymized output of the given input dataset which satisfies k-anonymity. Intrinsically, the notion of k-anonymity is very weak. Then classical k-anonymization based on k-anonymity is also vulnerable to re-identification attacks when some individuals have extreme values. For example, we assume that the input dataset contains the monthly income of individuals in a town. The adversary has known that only one individual's monthly income has been over $200\,K$ in this town. When $k(= 15)$-anonymization generates the output dataset which contains one group taking $[50\,K, 210\,K]$ as the generalized value of monthly income, the adversary can conclude the individual is in the group and the individual's monthly income is $210\,K$. The adversary can re-identify the individual with a probability that is over $1/15$.

Most classical k-anonymization that computes generalization values according to the input dataset is sensitive to extreme values, thus leaking private information.

3.2 (k, ρ)-Anonymization

Classical k-anonymization does not provide sufficient privacy preservation. We consider that it does not properly capture the notion of hiding in a k crowd,

because it does not impose any constraints on the mechanism used to generate generalized outputs, just as also mentioned in [14] and [7]. We aim to develop a new secure k-anonymization privacy definition whose mechanism used does not overly depend on an individual in the input dataset, i.e., it can achieve the notion of hiding in a k crowd.

In Definition 1, $I(t) \in \mathcal{I}_D$ can be denoted by $t \in D$. For convenience, we use $\mathbf{t}$ to denote $t \in D$, $M(D)$ to denote $M(D) = R$. Then Eq. 1 can be written $Pr[\mathbf{t}|M(D)] \leq \rho$.

Definition 2. *For all databases D, an individual $t \in D$ ρ-hides in a k crowd in D with respect to a mechanism M if $Pr[\boldsymbol{t}|M(D)] \leq \rho$ and $\rho \leq \frac{1}{k}$.*

Definition 3 *((k, ρ)-anonymization).* *A mechanism M is (k, ρ)-anonymization if for any database D and each individual $t \in D$, t ρ-hides in a k crowd in the database D.*

(k, ρ)-anonymization requires that for each individual $t \in D$, t hides in a k crowd in D. Individual t is indistinguishable from at least other $k - 1$ individuals by means of the mechanism M regardless of what the database D is, i.e., an adversary can re-identify t with a probability less than $\rho \leq \frac{1}{k}$. Thus the attribute value of the individual t can be changed to the value of any other individuals in the k crowd. The mechanism M does not release any re-identifying privacy information on the individual t except the common information in a k crowd.

Many mechanisms achieving k-anonymization generalize a value in the input database by means of replacing specific values with general values, such as replacing a specific monthly income with monthly income range. As described in Sect. 3.1, if it is not used carefully, the privacy information may be leaked. Most of these mechanisms do not satisfy (k, ρ)- anonymization. If the data can be generalized appropriately, it is possible to achieve (k, ρ)-anonymization.

(k, ρ)-anonymization is not sufficiently strong privacy preservation method in all scenarios. It is crucial for relaxing the notion of differential identifiability that an adversary may observe some common information on the individual t when the adversary knows every individual in a k crowd except the individual t. In a sense, this can be viewed as a privacy leak that is not allowed in differential identifiability. We consider that the leaked information on t is non-sensitive information, because it is shared by a k crowd. Such relaxation is needed in some scenarios that sacrifice non-sensitive information for improved utility while the individual t is not re-identified with a probability over $\frac{1}{k}$.

(k, ρ)-anonymization can be viewed as the relaxation of the notion of differential identifiability, thus there is a relationship between (k, ρ)-anonymization and differential identifiability.

Proposition 1. *A mechanism M satisfies ρ-differential identifiability. Then M is (k, ρ)-anonymization for $\rho \leq \frac{1}{k}$ (any integer $k \geq 2$) and any database D of size at least k.*

Proof. If the mechanism M satisfies ρ-differential identifiability, for any database D of size at least k and every individual $t \in D$, we can know

$$\begin{aligned} &Pr[I(t) \in \mathcal{I}_D | M(D) = R] \\ =& Pr[t \in D | M(D) = R] \\ \leq& \rho. \end{aligned}$$

According to Definition 1, t is re-identified in D with a maximum probability of $\rho(\leq \frac{1}{k})$, i.e.,

$$Pr[\mathbf{t}|M(D)] \leq \rho,$$

so the individual t can ρ-hide in a k crowd in the database D. The mechanism M is (k, ρ)-anonymization.

Proposition 2. *For any mechanism M, any database D of size at least k and each individual $t \in D$, the mechanism M satisfies ρ-differential identifiability if and only if for any integer $k \geq 2$ and $\rho = \frac{1}{k}$, t ρ-hides in a k crowd in the database D, i.e., M is (k, ρ)-anonymization.*

Proof. The "only if" direction can refer to Proof 3.2. If the mechanism M satisfies ρ-differential identifiability,

$$\begin{aligned} &Pr[I(t) \in \mathcal{I}_D | M(D) = R] \\ =& Pr[t \in D | M(D) = R] \\ =& Pr[\mathbf{t}|M(D)] \\ \leq& \rho = \frac{1}{k}. \end{aligned}$$

t can ρ-hide in a k crowd in the database D that satisfies (k, ρ)-anonymization.

For the "if" direction, the mechanism M is (k, ρ)-anonymization. The individual t ρ-hides in a k crowd in the database D. Assume that the mechanism M does not satisfy ρ-differential identifiability for the sake of contradiction, then there exists

$$\begin{aligned} &Pr[I(t) \in \mathcal{I}_D | M(D) = R] \\ =& Pr[t \in D | M(D) = R] \\ >& \rho. \end{aligned}$$

Then for any database D of size at least k and each individual $t \in D$, t is re-identified in D with a probability over $\rho = \frac{1}{k}$, i.e.,

$$Pr[\mathbf{t}|M(D)] > \rho.$$

Thus the individual t does not hide in a k crowd, which contradicts the fact that M is (k, ρ)-anonymization.

3.3 Privacy Axiom

Kifer and Lin [10] considered that the questions such as what makes a good privacy definition and how the data publisher should choose one must be addressed axiomatically. They presented the two axioms, namely the Privacy Axiom of Choice and the Transformation Invariance. The former allows us to randomly choose a privacy mechanism as long as this decision is not influenced by input database. It is a fundamental axiom which is required for any application of statistical privacy. The latter states that postprocessing sanitized data maintains privacy as long as the postprocessing mechanism does not deal with the sensitive information directly. We now show that the notion of (k, ρ)-anonymization also satisfies the two axioms.

Proposition 3. *Given two mechanisms M_1 and M_2 that both are (k, ρ)-anonymization, for any $p \in [0, 1]$, M_p is the mechanism that outputs M_1 with probability p and M_2 with probability $1-p$ on input database D, then M_p is also a privacy mechanism that satisfies (k, ρ)-anonymization.*

Proof. Since both M_1 and M_2 are (k, ρ)-anonymization, for any database D and each individual $t \in D$, we have

$$\begin{aligned} Pr[\mathbf{t}|M_p(D)] &= pPr[\mathbf{t}|M_1(D)] + (1-p)Pr[\mathbf{t}|M_2(D)] \\ &\leq p \cdot \rho + (1-p) \cdot \rho \\ &= \rho \\ &\leq \frac{1}{k}. \end{aligned}$$

Therefore, the mechanism M_p also is (k, ρ)-anonymization.

Proposition 4. *Let M_1 be (k, ρ)-anonymization. For a randomized M_2 whose input space is the output space of M_1, $M(\cdot) = M_2(M_1(\cdot))$ also is (k, ρ)-anonymization.*

Proof. The randomness in M_2 is independent of both the data and the randomness in the mechanism M_1. We have

$$\begin{aligned} & Pr[\mathbf{t}|M_2(M_1(D))] \\ =& Pr[t \in D|M_2(M_1(D)) = R] \\ =& \frac{Pr[D = D' \cup \{t\}] \cdot Pr[M_2(M_1(D' \cup \{t\})) = R]}{Pr[M_2(M_1(D)) = R]} \\ =& \frac{Pr[D = D' \cup \{t\}] \cdot Pr[M_1(D' \cup \{t\}) = S] Pr[M_2(S) = R]}{Pr[M_1(D) = S] Pr[M_2(S) = R]} \\ \leq& \rho \\ \leq& \frac{1}{k}. \end{aligned}$$

We know that $M(\cdot) = M_2(M_1(\cdot))$ also is (k, ρ)-anonymization.

3.4 Composition Axiom

(k, ρ)-anonymization can show that an adversary only know some common information shared by a k crowd, which can be viewed as the relaxation of differential identifiability. Thus (k, ρ)-anonymization takes advantage of the adversary's uncertainty about the input database, i.e., the definition does not assume the adversary's background knowledge about database. Inevitably, there exists some weaknesses. It has appeared in [6,7,14] that any privacy definition which exploits the adversary's uncertainty. Let M_1 and M_2 be (k, ρ)-anonymization. Since the two crowds generated by M_1 and M_2 for an individual t may be basically disjoint, the new crowd generated by the combination of M_1 and M_2 includes the individual t and can be very small. The adversary can re-identify individual t with a probability over $\frac{1}{k}$. Thus (k, ρ)-anonymization can release the output of a database in the non-interactive model only once.

Although (k, ρ)-anonymization mechanism is not robust under composition, we expect that it can compose (k, ρ_1)-anonymization mechanism with ρ_2-differential identifiability mechanism to obtain a (k, ρ)-anonymization mechanism, while ρ is a function of ρ_1 and ρ_2. Such composition can be useful in certain scenario. Given the anonymized output database that satisfies (k, ρ)-anonymization, one can release the database in the interactive model and meanwhile answer the queries that use the mechanisms satisfying ρ-differential identifiability.

Proposition 5. *Assume that mechanism M_1 is (k, ρ_1)-anonymization and M_2 is a mechanism that satisfies ρ_2-differential identifiability, then the mechanism $M = (M_1, M_2)$ is $(k, \rho_1\rho_2 m)$-anonymization.*

Proof. Let D be any database and t be any individual in D. The mechanism M_1 is (k, ρ_1)-anonymization making

$$\begin{aligned} Pr[\mathbf{t}|M_1(D)] &= Pr[t \in D|M_1(D)] \\ &\leq \rho_1. \end{aligned}$$

M_2 is a ρ_2-differential identifiability mechanism making

$$\begin{aligned} Pr[I(t) \in \mathcal{I}_D|M_2(D) = R] &= Pr[t \in D|M_2(D)] \\ &\leq \rho_2. \end{aligned}$$

We have

$$\begin{aligned} Pr[\mathbf{t}|M(D)] &= Pr[\mathbf{t}|(M_1(D), M_2(D))] \\ &= Pr[t \in D|(M_1(D), M_2(D))] \\ &= \frac{Pr[D = D' \cup t]Pr[M_1(D' \cup t), M_2(D' \cup t)]}{Pr[M_1(D), M_2(D)]} \\ &= \frac{Pr[D = D' \cup t]Pr[M_1(D' \cup t)]Pr[M_2(D' \cup t)]}{Pr[M_1(D)]Pr[M_2(D)]} \\ &= \frac{Pr[D = D' \cup t]Pr[M_1(D' \cup t)]}{Pr[M_1(D)]} \cdot \frac{Pr[M_2(D' \cup t)]}{Pr[M_2(D)]} \\ &\leq \rho_1\rho_2 m. \end{aligned}$$

When the values of ρ_1 and ρ_2 are taken carefully, it can make $\rho_1\rho_2 m \leq \frac{1}{k}$. Thus the individual t can $\rho_1\rho_2 m$-hides in a k crowd with respect to M in the database D.

It is normal that privacy level degrades as more information is leaked. As mentioned in the proof, we must control the values of ρ_1 and ρ_2 in a good way to satisfy $(k, \rho_1\rho_2 m)$-anonymization. Note that when the value of m is too large, the values of ρ_1 and ρ_2 will become very small to satisfy $\rho_1\rho_2 m \leq \frac{1}{k}$ which introduces too much noise. Furthermore, an adversary may conclude with high confidence that the individual t is not in the database D after observing query results and considering the values of ρ_1 and ρ_2 are very small. It is evident that the composition between differential identifiability mechanism and other weaker mechanism should be used in a well-controlled way.

4 Conclusions

In this paper, we proved the properties of differential identifiability under composition that can allow us to apply differential identifiability for complex privacy queries in data mining and data publishing. We identified the weaknesses of the k-anonymization methods and provided the notion of (k, ρ)-anonymization. (k, ρ)-anonymization can avoid the vulnerabilities existing in classical k-anonymization and achieve the relaxation of the notion of differential identifiability. We also studied the power and the potential weaknesses under composition. The results show that it is important to control the privacy parameters to prevent privacy information leaked when (k, ρ)-anonymization is used in the interactive model. Our achievements provide basis for practical application of differential identifiability and will facilitate the development of practical privacy-preserving data mining algorithms supporting differential identifiability.

Acknowledgment. This project was supported by the National Key Research and Development Program of China (No. 2016YFC1000307) and the National Natural Science Foundation of China (No. 61571024, 61971021) for valuable helps.

References

1. Blum, A., Dwork, C., McSherry, F., Nissim, K.: Practical privacy: the SuLQ framework. In: Proceedings of the Twenty-Fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2005, pp. 128–138. ACM Press, New York (2005)
2. Dinur, I., Nissim, K.: Revealing information while preserving privacy. In: Proceedings of the Twenty-Second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2003, pp. 202–210. ACM Press, New York (2003)
3. Dwork, C.: Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4052, pp. 1–12. Springer, Heidelberg (2006). https://doi.org/10.1007/11787006_1

4. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
5. Dwork, C., Nissim, K.: Privacy-preserving datamining on vertically partitioned databases. In: Franklin, M. (ed.) CRYPTO 2004. LNCS, vol. 3152, pp. 528–544. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-28628-8_32
6. Ganta, S.R., Kasiviswanathan, S.P., Smith, A.: Composition attacks and auxiliary information in data privacy. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2008, pp. 265–273. ACM Press, New York (2008)
7. Gehrke, J., Hay, M., Lui, E., Pass, R.: Crowd-blending privacy. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO 2012. LNCS, vol. 7417, pp. 479–496. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32009-5_28
8. Gehrke, J., Lui, E., Pass, R.: Towards privacy for social networks: a zero-knowledge based definition of privacy. In: Ishai, Y. (ed.) TCC 2011. LNCS, vol. 6597, pp. 432–449. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-19571-6_26
9. He, Y., Naughton, J.F.: Anonymization of set-valued data via top-down, local generalization. Proc. VLDB Endow. **2**(1), 934–945 (2009)
10. Kifer, D., Lin, B.R.: Towards an axiomatization of statistical privacy and utility. In: Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, pp. 147–158. ACM Press, New York (2010)
11. Kifer, D., Machanavajjhala, A.: No free lunch in data privacy. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, pp. 193–204. ACM Press, New York (2011)
12. Lee, J., Clifton, C.: Differential identifiability. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, KDD 2012, pp. 1041–1049. ACM Press, New York (2012)
13. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: privacy beyond k-anonymity and l-diversity. In: 22nd International Conference on Data Engineering, ICDE 2007, pp. 106–115. IEEE Computer Society Press, Los Alamitos (2007)
14. Li, N., Qardaji, W., Su, D.: On sampling, anonymization, and differential privacy or, k-anonymization meets differential privacy. In: Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security, ASIACCS 2012, pp. 32–33. ACM Press, New York (2012)
15. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-diversity: privacy beyond k-anonymity. In: 22nd International Conference on Data Engineering, ICDE 2006, p. 24. IEEE Computer Society Press, Los Alamitos (2006)
16. Mendes, R., Vilela, J.P.: Privacy-preserving data mining: methods, metircs, and applications. IEEE Access **5**, 10562–10582 (2017)
17. Samarati, P.: Protecting respondents identities in microdata release. IEEE Trans. Knowl. Data Eng. **13**(6), 1010–1027 (2001)
18. Sweeny, L.: Achieving k-anonymity privacy protection using generalization and suppression. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **10**(5), 571–588 (2002)
19. Sweeny, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertain., Fuzziness Knowl.-Based Syst. **10**(5), 557–570 (2002)
20. Yu, S.: Big privacy: challenges and opportunities of privacy study in the age of big data. IEEE Access **4**, 2169–3536 (2016)