International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-6, Issue-2, March 2018
DOI:10.21276/ijircst.2018.6.2.3

Privacy Risk Against Composition Attack

A H M Sarowar Sattar, Sumyea Helal

Abstract— Privacy in multiple independent data publishing has attracted considerable research interest in recent years. Although each published data set poses a small privacy risk to individuals, recent studies show that this risk increases when different organizations have some common records and they publish their data sets independently without any coordination with each other. If an individual can be detected from disparate providers, the individual's privacy is compromised. This type of privacy breach is called composition attack. A few studies have done to mitigate this attack. However, none of them studies the risk against this attack from a single data set. Motivated by this gap, this paper uses a probabilistic model to estimate the risk against composition attack from a single data. Therefore, a publisher can predict the risk against composition attack of a data set prior to publication. To evaluate the effectiveness of our model we also perform empirical analysis to show that the estimated risk can give us the pattern of the real risk.

***Index Terms*— Privacy, Composition attack, Anonymization.**

I. INTRODUCTION

The problem of privacy preserving data publishing has received a lot of attention in recent years. Data anonymization is one type of privacy preserving data publishing method that seeks to hide the identity of a record owner. The assumption is that though the identities of the data owners are hidden, still it is possible to perform data analysis on the sensitive information. All the multiple data publishing techniques are restricted to single publisher [12,13,20,23] and do not support data publication from overlapping records from multiple independent publishers [6,21,24,25]. Note that we use the word publication and publishing interchangeably. The situation of multiple independent data publishing is different than other data publishing techniques.

Manuscript Received March 15, 2018.

A H M Sarowar Sattar, Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh, Mobile No. 8801761325290, (e-mail: sarowar@gamil.com)

Sumyea Helal, Department of Computer Science and Engineering, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh. (e-mail: sumyeahelal@gmail.com)

Here the assumptions are; one publisher has no knowledge about the overlapping individuals' record with other publishers' data set and how other publishers are going to anonymize those records of overlapping individuals. Due to overlapping records, an adversary can collect different releases from different publishers and then perform intersection among those releases to reveal sensitive information of an individual. This type of attack is reported in [7] and known as composition attack. In this attack, an adversary follows the sensitive attribute to reveal sensitive information of a victim without precisely identifying victim's record.

Table 1: Original data set of (a) Hospital-A (b) Hospital-B

Name	Age	Sex	Zip Code	Diagnosis
Emu	25	m	5095	A
Alex	24	m	5085	B
Clark	20	m	5001	C
Hafiz	23	m	5005	D
Alice	22	f	5095	E
Mina	25	f	5201	F
Sofia	20	f	5002	B
Anju	21	f	5087	F
Bob	38	m	5195	B
Lalin	35	m	5087	D
Thuc	28	m	5102	G
Abed	31	m	5202	F

(a)

Name	Age	Sex	Zip Code	Diagnosis
Hussy	25	m	5095	H
Michel	24	m	5085	H
Bokul	20	m	5031	C
Safiq	23	m	5025	C
Alice	22	f	5095	E
Lima	25	f	5065	D
Nima	20	f	5202	B
Fami	21	f	5177	C
Bob	38	m	5195	B
Lalin	35	m	5187	E
Kona	28	f	5042	D
Bipa	31	f	5032	F

(b)

Table 2: Anonymous data set of (a) Hospital-A (b) Hospital-B

Age	Sex	Zip Code	Diagnosis
15-25	m	50**	A
15-25	m	50**	B
15-25	m	50**	C
15-25	m	50**	D
15-25	f	5***	E
15-25	f	5***	F
15-25	f	5***	B
15-25	f	5***	C
26-40	m	5***	B
26-40	m	5***	D
26-40	m	5***	G
26-40	m	5***	F

(a)

Age	Sex	Zip Code	Diagnosis
10-30	*	50**	H
10-30	*	50**	H
10-30	*	50**	C
10-30	*	50**	C
10-30	*	50**	E
10-30	*	50**	D
31-50	*	5***	B
31-50	*	5***	C
31-50	*	5***	B
31-50	*	5***	E
31-50	*	5***	D
31-50	*	5***	F

(b)

To illustrate the problem, let us assume that Hospital-A and Hospital-B releases Table-2(a) and Table-2(b) of the original Table-1(a) and Table-1(b), respectively. Alice's (22 years old female living in an area of zip code 5095) equivalence class (the groups of data in the published data set, that match with Alice's record) in both tables has only one common sensitive value; i.e., **E**. So, an adversary knowing the non-sensitive information (i.e. age, sex, zip code) of Alice and the fact that she has visited two hospitals can derive her sensitive value from both tables. Individually both anonymous data sets pose low privacy risk but collectively compromise the privacy of overlapping patients

due to the composition attack. In other words, under the composition attack the independent anonymous releases cannot retain privacy.

Some anonymization techniques are vulnerable to composition attack. Anonymization techniques [3,4,8,9] are widely discussed in literature and well known schemes includes partitioning based approach [11,13,14,15,29] (achieved either by generalization or suppression) and perturbation based approach [5,17,18,26] (mostly achieved by adding random noise). Basically, the techniques based on partition based approach are vulnerable to composition attack. Because, it publishes the sensitive values in its original form and it is possible to locate the small group where the individual's record resides. In this paper, we consider only the partition based schemas where generalization is applied for data anonymization.

A little studies have done on composition attack. A few methods [1,2,7] have been designed to deal with composition attack, but none of them can handle it completely. Study in [7] shows that existing partition based anonymization approach cannot prevent individual privacy for this attack. Another study also shows that there is not a general solution to prevent the composition attack in a generalization-only publication scheme [1]. However, so far our knowledge no work has done to study risk associated with an anonymous data set. If it is possible to quantify that risk before publishing the data set, it may help the publishers to take action accordingly and that may reduce the privacy disclosure. Therefore, in this paper we will try to estimate the risk associated with an anonymous data set prior to publication.

It is challenging to estimate the risk of composition attack when another data set is not available to a publisher or a publisher does not have knowledge about the overlapping individuals. A simulation of an anonymous data set from other publishers can be a solution here. If there is a such simulated data set then a publisher can use that (simulated data set) to estimate the risk associated with an anonymous version against composition attack. Therefore, the problem of estimating the risk of composition attack becomes a problem of simulating a data set from other publishers.

Simulation of such a data set can be done by using the global statistics of a population from where the individuals' records are collected. Generally speaking, overlapping records occur in different publishers' data sets when those publishers are located at the same location or deal with individuals from the same population. In such case, we can assume that the collected data/records by different publishers should have more or less similar statistical information, as each data set is sampled from a common underline population. Therefore, a publisher can use the global statistics of that population or in worst case, can use the statistical information from his/her collected data set to simulate an anonymous data set from other publishers. Finally, the contributions of this paper can be summarized as follows:

- So far our knowledge, this is the first work to deal with the risk associated with an anonymous data set against composition attack.
- We have shown a way to simulate an anonymous data set from other publishers where all publishers collect their data sets from a common underline population.

II. PRELIMINARIES

Let D={t_1,t_2,….,t_n} be a multi-set of records, where each record t_i represents the information of an individual i. Each record t_i = {id_i, q_i, s_i}, where id_i ∈ ID, q_i ∈ QID, and s_i ∈ S. ID represents unique identifiers, which is used to uniquely identify a record such as name or medicare card number. QID is a set of other attributes that can potentially identify a person, such as age, zip code and sex, and S is the set of sensitive value such as disease. The quasi-identifiers QID={q_1,q_2,…,q_m} consist of m attributes, each of which is associated with an attribute taxonomy.In a published data set, the attribute ID has been removed, QID attributes and sensitive attributes are kept in the published data sets. An adversary may use record linkage [22] between QID attributes and external information to link an individual's identity to their sensitive information. To avoid this disclosure, one frequently used solution is to replace the QID values with more general values from its taxonomy, so that the individuals in an equivalence class are indistinguishable and their sensitive values cannot be inferred with a high confidence. Some well known principles are k-anonymity [20], l-diversity [13], (alpha, k)-anonymity [23] and t-closeness [12]. Therefore, consider D^*={t`_1},t`_2,…,t`_n} be a published data set, where t`=q`_1,q`_2,…,q`_m,s} and q`_i is any value from the taxonomy of q_i.

Definition 1 (Equivalence Group) For an anonymous data set, an equivalence group is a multi-set of records/tuples in that data set having identical values in QID attributes.

For example, tuples 1 to 4 in Table 2(a) form an equivalence group with respect to {age, sex, zip code}, because their corresponding values are identical.

For simplicity of discussion, let us consider there are two independent anonymous data set D_1^* and D_2^*. We use the notation $E_i(D_j^*)$ to represent the equivalence group of an individual i, $QI(E_i(D_j^*))$ is the set of all values in the quasi-identifiers, and use the notation $S(E_i(D_j^*))$ to represent the multi set of sensitive value in $E_i(D_j^*)$ of a published data set D_j^*.

Definition 2 (Match [27]) Let $E_i(D_1^*)$ and $E_i(D_2^*)$ be two equivalence groups in D_1^* and D_2^*, respectively. $E_i(D_1^*)$ and $E_i(D_2^*)$ match if every value pair in $QI(E_i(D_1^*))$ and $QI(E_i(D_2^*))$ are equal or have a non-empty intersection, and $S(E_i(D_1^*)) \cap S(E_i(D_2^*)) \neq$ NULL.

For example, let $E_i(D_1^*)$ and $E_i(D_2^*)$ be two equivalence classes where $QI(E_i(D_1^*))$ =(30-40, M, 5000-5100), $QI(E_i(D_2^*))$=(20-40, M, 5050-5100), $S(E_i(D_1^*))$ = (A, B, B, C) and $S(E_i(D_2^*))$ = (C, B, B, E). Then $E_i(D_1^*)$ and $E_i(D_2^*)$ are called match.

Privacy of an individual in an equivalence group depends on the result of intersection of the match equivalence groups, and the number of distinct sensitive values measures the anonymity factor of an individual in that equivalence group. In the above example,

distinct($S(E_i(D_1^*)) \cap S(E_i(D_2^*))$ = { B, C}, therefore, sensitive attribute of individuals in those equivalence

groups can be mapped with two different sensitive values; B and C.Therefore, the privacy breach of an individual is formally defined by Definition 3.

Definition 3 (Privacy Breach) Given published data sets D_1* and D_2* and the knowledge that a victim v has records in both published data sets. A privacy breach occurs if the result of the intersection of the match equivalence groups of an individual i=v is less than the value (say l) defined by the publishers. Therefore, the privacy breach occurs when, distinct($|S(E_i(D_1*)) \cap S(E_i(D_2*))|$) < l, where l represents a publisher's predefined protection parameter.

For instance, let us consider l=2, then Alice's privacy is breached in anonymous releases of Tables 2(a) and (b). Because, *distinct*($|S(E_{Alice}(D_1*)) \cap S(E_{Alice}(D_2*))|$) < 2, an adversary can identify that Alice is suffering from E.

To estimate the risk of an anonymous data set, we need to look for those equivalence groups which are under the threat of composition attack, instead each individual's privacy breach. Because, only the overlapping individuals are subject to composition attack and a publisher does not have knowledge or information regarding those overlapping records. A vulnerable equivalence group of an anonymous data set is defined by Definition 4.

Definition 4 (Vulnerable equivalence group) Given an equivalence group $E_i(D_j*)$ of an anonymous data set D_j*. The equivalence group $E_i(D_j*)$ is considered as vulnerable equivalence group against composition attack when privacy breach (Definition 3) occurs for any individual i in that equivalence group. Therefore, the set of vulnerable equivalence group,

$$vul(E_i(D_j^*)) = \exists_i\{E_i(D_j^*) \in D_j^* : distinct(|S(E_i(D_1^*)) \cap S(E_i(D_2^*))|) < l\}$$

Now, given a publisher predefine protection parameter l for an anonymous data set D_1*, we denote by $|E(D_1*)|$ the set of all equivalence group in D_1*. We measure the risk of D_1* against composition attack as the percentage of equivalence groups under the threat of that attack.

$$Risk of(D_1^*) = \frac{|vul(E_i(D_1^*))|}{|E_i(D_1^*)|} \times 100\% \qquad (1)$$

where $|vul(E(D_1*))|$ represents the total number of vulnerable equivalence group in D_1*.

III. RISK ESTIMATION MODEL

In this section, we use the privacy model that was first proposed in [28] by Sarowar and et. al to anonymized a dataset to protect from composition attack. Let us consider that the original data sets D_1 and D_2 are samples from the large population Ω and the intersection of D_1 and D_2 are not empty. D_1* and D_2* are the anonymous version of them respectively. Another data set D_0* is a hypothesized data set of D_2*. All the data sets have same attribute domain. For simplicity of discussion, we further assume that the size of all data sets are same. We use the hypothesized data set D_0* to simulate D_2*. Consider, D_0* is a random sample of Ω with record probability (Definition 5) P(t`), where t` is a record with sensitive value s.

Definition 5 (Record probability [29]) We assume that attribute values and the sensitive value in a record are independent. $P(q_i)$ and P(s) are the frequencies of value q_i and sensitive value s in the population. The probability of a record t` = {$q_1$`, q_2` …, $q_m$`, s}, denoted as P(t`), is assigned as the following.

$$P(t`) = P(q_1`) \times P(q_2`) \dots\dots\dots \times P(q_m`) \times P(s)$$
$$= \left(\prod_{i=1}^{m} P(q_1`) \times P(s)\right) \qquad (2)$$

For example, let us assume that P(15-30) = 0.15, P(male) = 0.5, and P(C) = 0.05 are obtained from the patient population. Let t` = {15-30, male, diabetes}. P(t`) = 0.00375. Note that the publisher's knowledge may include that a 40-60 male has higher probability of diabetes, say 0.02. Such knowledge can also be modeled. The independency assumption is used when we do not have other knowledge.

Definition 6 (Hypothesized data set) A hypothesized data set D_0* is a data set which is created by using n random draws from the global population Ω, where n is the size of that data set and record probability P(t) of a record t represents its chance to be appeared in D_0* in a random draw. Therefore,

D_0* = Collection of n random draws with record probability.

According to [28], $P_{D0*}(E_i(D_0*), s^l)$ represent the probability that l different sensitive values will be common in equivalence group $E_i(D_0*)$ and $E_i(D_1*)$. Therefore, the vulnerable equivalence class can be defined as follows to calculate the risk of a data set against composition attack.

Definition 7 (Vulnerable equivalence group 2) Given published data sets D_1* and a hypothesized data set D_0*, and the confidence level C of a publisher. An equivalence group $E_i(D_1*)$ is vulnerable against composition attack when the probability of appearing l common sensitive values in $E_i(D_1*)$ is less than publisher's confidence level C. Therefore, those equivalence groups are considered as vulnerable equivalence groups for which $\forall s^l (P_{D0*}(E_i(D_0*), s^l) < C$, where l represents a publisher's predefined protection parameter, and $E_i(D_1*)$ and $E_i(D_0*)$ are the equivalence groups of an individual i in D_1* and D_0*, respectively.Based on this framework, we can define the objective of this work. Given a data set D_1*, has already been anonymized, the expected number of shared sensitive value l and publisher's confidence parameter C our objective is to find out percentage of equivalence groups that are vulnerable to composition attack when another similar anonymous data set will be available to an adversary.

IV. EXPERIMENTAL RESULT

In this section we describe our experimental study. The primary goal is to quantify the risk against composition attack of an anonymous data set. Although the earlier works study the severity of composition attack using both anonymous data sets (from different publishers who have records from overlapping individuals), to best of our knowledge, none of these works study the risk from a single anonymous data set. At the end of this section, we also present the result by using multiple anonymous data sets to validate our estimation. We use one of partition based anonymization techniques to demonstrate the risk of a composition attack: k-anonymity. For implementing

k-anonymity we use the Mondrian multidimensional approach proposed in [11].

Table 3: Attribute domain size

Attribute	Age	Sex	Education	Race	Birth Place	Occupation	Salary
Domain Size	100	2	20	6	41	50	50

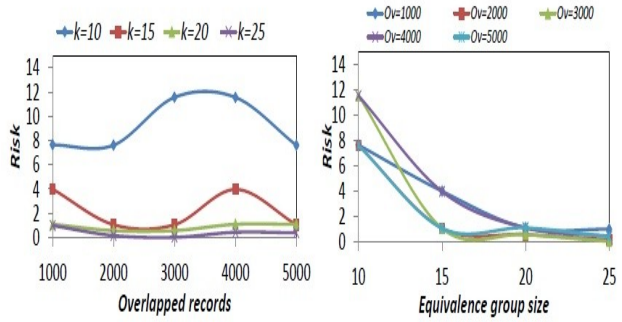

Figure-1: Estimated privacy risk against composition attack for different equivalence group size and overlapping records.

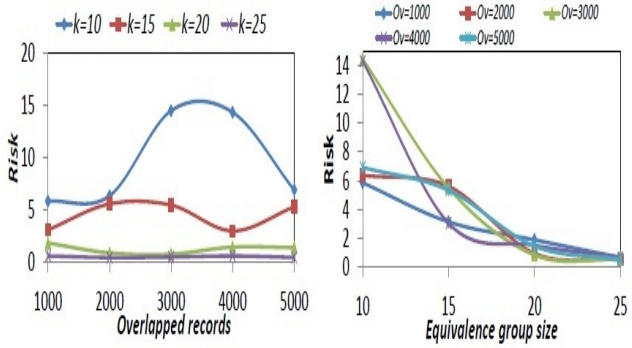

Figure-2: Real privacy risk against composition attack for different equivalence group size and overlapping records.

We performed experiments with real world data sets from the U.S. census Bureau (available at http://ipums.org). This data set consists of 600k tuples. From this data set we use five attributes: age, sex, education, race and birth place as quasi identifying attributes and the occupation attribute as the sensitive attribute. All QI attributes are discrete except age and education. The sizes of their domains are reported in Table 3.

We composed two disjoint data sets from that data set via random draws of 100,000 tuples and considered them as two independent group of data sets, say D_1 and D_2, from different publishers. The remaining 300k tuples are used as an overlapping pool. We made five copies of each data set in each group and randomly inserted 1k, 2k, 3k, 4k and 5k tuples from the overlapping pool to the copies respectively yielding 5 sets of data of size 101k, 102k, 103k, 104k and 105k. We inserted those overlapping pools in two groups in such a way that each set of data sets of same size share overlapping tuples of 1k, 2k, 3k, 4k and 5k respectively.

To estimate the risk of an anonymous data set against composition attack, in this experiment we use the data sets from one group of different sizes. All the data sets are anonymized by using Mondrain algorithm proposed in [11] and estimated privacy risk is measured on those anonymous data sets. Vulnerable equivalence groups are considered using Definition 7. Later on the risk on an anonymous data set is estimated by using Equation 1.

Figure-1 shows the result of that estimated privacy. Figure-1(a) is plotted with respect to different overlapping records for different equivalence group size and Figure-1(b) is plotted in the other way around. In Figure-1(a), there is no fixed pattern of that privacy risk. Sometime it (privacy risk) increases and sometime decreases with increase in overlapping records and all other experiments (for different

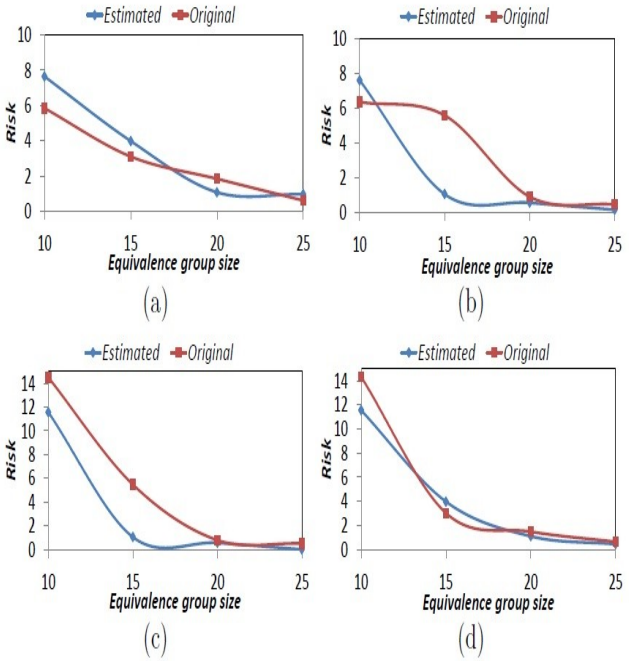

Figure-3: Comparison with estimated and real privacy risk for different overlapping records (a) 1000 (b)2000 (c)3000 and (d) 4000

equivalence group size) also experience the same. We believe this is true. Because, when overlapping individual increase, the overlapping record can increase in any pattern (either the overlapped records in same group or different group). Therefore, if overlapping records are increasing in same equivalence group, which means in that group we have more common sensitive values, thus less chance of a successful composition attack.To measure the real privacy risk, we anonymize the both groups of data set and then composition attacks are conducted between all pairs of data sets with the same overlapping tuples. Vulnerable equivalence groups are considered using Definition 4. Later on the risk on an anonymous data set is estimated by using Equation 1. Figure-2(a) and Figure-2(b) show the practical risk found with respect to different overlapping records and equivalence group sizes respectively. Here, we see the risk observed here are almost have same pattern as we have estimated from a single data set.

Figure-3 shows the comparison of estimated and real privacy risk against composition attack. In Figure-3(a) and Figure-3(d) the estimated and the real privacy risks are very

similar. But, in Figure-3(b) and Figure-3(c), most of the time the estimated risk is less than the real one. However, in all cases the estimated risk can capture the risk pattern associated with an anonymous data set.

V. CONCLUSION

This paper presents a model to estimate the risk associated with an anonymous data set prior to publication. We have provided a theoretical foundation to measure the risk. Furthermore, we have experimentally shown that our risk estimation model can capture the pattern of the risk against composition attack associated with an anonymous data set. In this work, we consider that all attributes of a record are independent. However, a sensitive attribute such as disease is not independent from the non-sensitive attribute like age, sex and others. For example, female patient have higher chance to get breast cancer than male patient. However, we believe this assumption is a good starting point for a new approach to mitigate composition attack in multiple independent publication.

REFERENCES

[1] Muzammil M Baig, Jiuyong Li, Jixue Liu, and Hua Wang. Cloning for Privacy Protection in Multiple Independent Data Publications. In CIKM, pages 885-894,2011.

[2] Muzammil M. Baig, Jiuyong Li, Hua Wang, and Jixue Liu. Studying genotype-phenotype attack on k-anonymised medical and genomic data. In CRPIT, pages159-166, 2009.

[3] Bee-chung Chen, Kristen Lefevre, and Raghu Ramakrishnan. Privacy Skyline :Privacy with Multidimensional Adversarial Knowledge. In VLDB, pages 770-781, Vienna, Austria, 2007. ACM.

[4] Bee-Chung Chen, Kristen Lefevre, and Raghu Ramakrishnan. Adversarial-knowledge dimensions in data privacy. The VLDB Journal, 18(2):429-467, April,2009.

[5] Cynthia Dwork. Differential Privacy. In ICALP, pages 1-12. Springer, bugliesi,edition, 2006.

[6] Benjamin C. M. Fung, Ke Wang, Ada Wai-Chee Fu, and Jian Pei. Anonymity for continuous data publishing. In EDBT '08, pages 264-275, New York, New York,USA, 2008.

[7] Srivatsava Ranjit Ganta, Shiva Prasad, and Adam Smith. Composition Attacks and Auxiliary Information in Data. In SIGKDD, pages 265-273, 2008.

[8] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. Boosting the Accuracy of Differentially Private Histograms Through Consistency. Proceedings of the VLDB Endowment, 3(1):1021-1032, 2010.

[9] Wei Jiang and Chris Clifton. A secure distributed framework for achieving k-anonymity. The VLDB Journal, 15(4):316-333, August 2006.

[10] Daniel Kifer and Ashwin Machanavajjhala. A rigorous and customizable framework for privacy. In PODS '12, pages 77-88, New York, NY, USA, 2012.

[11]K. LeFevre, D.J. DeWitt, and R. Ramakrishnan. Mondrian Multidimensional K-Anonymity. In ICDE, pages 25-25, 2006.

[12]Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-Closeness : Privacy Beyond k-Anonymity and l-Diversity. In ICDE, number 3, pages 106-115, 2007.

[13]Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-Diversity: Privacy Beyond k-Anonymity. ACM Trans. on

Knowledge Discovery from Data, 1(1):3es, March 2007.

[14]B Malin. k-Unlinkability: A privacy protection model for distributed data. Data & Knowledge Engineering, 64(1):294-311, January 2008.

[15]Bradley Malin. Secure construction of k-unlinkable patient records from distributed providers. Artif. Intell. Med., 48(1):29-41, January 2010.

[16]Bradley Malin, Edoardo Airoldi, Samuel Edoho-eket, and Yiheng Li. Con_gurable Security Protocols for Multi-party Data Analysis with Malicious Participants. In

ICDE, number September, pages 533-544, 2004.

[17]David J Martin, Daniel Kifer, Ashwin Machanavajjhala, Johannes Gehrke, and Joseph Y Halpern. Worst-Case Background Knowledge for Privacy-Preserving Data Publishing. In ICDE, pages 126-135, 2007.

[18]Noman Mohammed, Rui Chen, Benjamin C M Fung, and Philip S Yu. Differentially Private Data Release for Data Mining. In SIGKDD, pages 493-501, 2011.

[19]M. Ercan Nergiz, Maurizio Atzori, and ChristopherW. Clifton. Hiding the presence of individuals from shared databases. In SIGMOD, pages 665-676, New York, New

York, USA, 2007.

[20]Latanya Sweeney. k-anonymity: A model for protecting privacy. Int'l Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5):1-14, 2002.

[21]Ke Wang and C.M. Benjamin Fung. Anonymizing Sequential Releases . In ACM SIGKDD, pages 414-423, 2006.

[22]W. Winkler. Advanced methods for record linkage. In Proceedings of the Selection on Survey Research Methods, Americal Statistical Society, pages 467-472, 1994.

[23]Raymond Chi-wing Wong, Jiuyong Li, Ada Wai-chee Fu, and Ke Wang. (Alpha,k)-Anonymity : An Enhanced k-Anonymity Model for Privacy-Preserving Data Publishing. In ACM SIGKDD, pages 754-759, 2006.

[24]R.C.-W. Wong, A.W.-C. Fu, Jia Liu, Ke Wang, and Yabo Xu. Global privacy guarantee in serial data publishing. In ICDE, pages 956-959, march 2010.

[25]Xiaokui Xiao and Yufei Tao. m-Invariance : Towards Privacy Preserving Re-publication of Dynamic Datasets. In SIGMOD, pages 689-700, 2007.

[26]Xiaokui Xiao, Guozhang Wang, Johannes Gehrke, and Thomas Je_erson. Differential Privacy via Wavelet Transforms. IEEE Trans. on Knowledge and Data

Engineering, 23(8):1200-1214, 2011.

[27] J. Li, M. M. Baig, A. S. Sattar, X. Ding, J. Liu, M. W. Vincent, A hybrid approach to prevent composition attacks for independent data releases, Information Sciences 367-368 (2016).

[28] A. S. Sattar, J. Li, J. Liu, R. Heatherly, B. Malin, A probabilistic approach to mitigate composition attacks on privacy in non-coordinated environments, Knowledge-Based Systems 67 (2014) 361- 372.

[29] A. S. Sattar, J. Li, X. Ding, J. Liu, M. Vincent, A general framework for privacy preserving data publishing, Knowledge-Based Systems 54 (2013) 276-287.

