A Survey on Privacy Preserving Methods against Composition Attack in Non-Coordinate System

Tamanna Rahaman
Department of Computer Science & Engineering
Bangladesh Army University of Engineering & Technology
Rajshahi, Bangladesh

Md. Muktar Hossain
Department of Computer Science & Engineering
Rajshahi University of Engineering & Technology
Rajshahi, Bangladesh

Fahmida Haque Mim
Department of Computer Science & Engineering
Bangladesh Army University of Engineering & Technology
Rajshahi, Bangladesh

ABSTRACT
Anonymizing data for safe publication has become a growing research field in recent years. Most of the earlier work in this field focused on independent publication. But these acquiring privacy methods fall short when considering the real-life situation of multiple releases of data by different organization. Numerous independent publications can contain information about the same person, and joining these separate publications can disclose that person's identity. This is called a composition attack. Several strategies have been developed to solve this problem and still it is in developing phase. Our paper will survey the most prominent methods proposed to anonymize common data of multiple independent publications. We hope to help the researchers come up with new ideas to mitigate the risk of a composition attack in non-coordinated system by summarizing the work that has already been done up to this point.

General Terms
Summery Paper, Pattern Recognition, Data Security, Algorithms

Keywords
Composition attack, Non-coordinate system, Independent publishing, Anonymization

1. INTRODUCTION
In the past thirty years, the growth of computer usage has pushed us into a new era of digitalization, where data has become the most important commodity. While the process has improved our existence, it has also put us at significant risk of privacy. Many of the apps and software we use to collect user data can be sensitive and a breach of personal privacy if widely known. The data owners sometimes need to publish these data for academic research or commercial purposes, like- hospitals, government facilities, insurance companies, or even social media companies. But before releasing this information, they have to make sure that any sensitive data cannot be traced back to one individual. The need for this data security has revealed the up latest and relatively uncharted field of study.

While much progress has been made in preserving privacy [12][14] through *k*-anonymity [1][4], *l*-diversity [3], m-invariance [7] and *t*-closeness [8], but all these techniques apply to one time publications [16]. There has been some study of serial publication [10], but the same organization publishes data with the same individual.

In that case, the publishers have previous knowledge about the different data sets. But if we consider the real-life application, we find that multiple organizations that release anonymized data separately have overlapping data about the similar people. Every organization can't have foreknowledge about all data sets published before them. If an adversary puts together these overlapping data, the security is greatly compromised. This is called a composition attack. Composition attacks expose a sensible and important class of susceptibility. The easiest way to mitigate against this attack is to coordinate between publishers, but that is hardly feasible with so many organizations in existence. As privacy-preserving data publishing becomes more commonly deployed, it is almost impossible to keep track of all the organizations that publish anonymized data about any individual or entity of a non-coordinate system. Thus using schemes that are vulnerable to composition attacks has become a potential for security risk.

In that paper we summarize the security measures taken against composition attack in non-coordinate system and also investigate the effectiveness of the discovered methods along with future scope of further development.

2. COMPOSITION ATTACK
Before we get into the methods to protect against composition attacks, we need to understand what it is and how it works [11]. As an example, we can consider someone who has visited more than one hospital for treatment. If both of these hospitals release anonymized data independently, an adversary who has this fore knowledge can put together the two data set to achieve the sensitive data of the victim. We have to consider that anyone trying to harm an individual will gather background knowledge [9] about the victim. It is explained better with an example below.

Example 1. Suppose two hospitals H1 and H2, in the same area release anonymized medical information about patients admitted into the hospital. Being in the same city, some patients may visit both hospitals with the same medical problem to get a second opinion. Tables 1(a) and 1(b) are independent medical data from H1 and H2 hospitals. Both tables are anonymized using k-anonymity, where the value of k for H1 is k =4 and for H2 is k =6. The patient's medical condition is the sensitive information that we need to protect. The other attributes are quasi-identifiers, and they are generalized so that within each group of rows, the vectors of non-sensitive attributes are the same. If an adversary of Alice knows that she is 26 years old, lives in zip code 13010, and recently visited both hospitals, then they can infer her location from joining both anonymized tables. Since both the tables are anonymized, Alice matches four records in H1's data and six

records in H2's. However, AIDS is the only disease that appeared in both lists, and so it can be concluded that Alice has AIDS. Thus Alice's sensitive information is exposed, and privacy breached.

Based on this, anyone with some background knowledge can narrow down the number of possible sensitive values for an individual by intersecting the sets of sensitive values present in his/her groups from multiple anonymized publications. If there are only two possibilities to choose from, the adversary can still guess 50% surety. Even if the adversary can narrow it down to a few options, there is an excellent chance of guessing right about the victim. The more possibilities there are, the bigger chance of the adversary guessing wrong, and the privacy will increase.

So, we can understand that a composition attack greatly threatens to destroy the privacy of published data set in case of multiple independent publications. This creates a great field of research as everyone wants to protect their privacy. Many methods have already been proposed over the years to overcome this weakness, and we will discuss the more prominent methods in this paper.

Table 1(a): Medical data of Hospital 1

	Non-Sensitive			Sensitive
	Zip code	**Age**	**Nationality**	**Condition**
1	130**	<35	*	AIDS
2	130**	<35	*	Tuberculosis
3	130**	<35	*	Flu
4	130**	<35	*	Tuberculosis
5	130**	<35	*	Cancer
6	130**	<35	*	Cancer
7	130**	≥35	*	Cancer
8	130**	≥35	*	Cancer
9	130**	≥35	*	Cancer
10	130**	≥35	*	Tuberculosis
11	130**	≥35	*	Viral Infection
12	130**	≥35	*	Viral Infection

Table 1(b): Medical data of Hospital 2

	Non-Sensitive			Sensitive
	Zip code	**Age**	**Nationality**	**Condition**
1	130**	<30	*	AIDS
2	130**	<30	*	Heart Disease
3	130**	<30	*	Viral Infection
4	130**	<30	*	Viral Infection
5	130**	≥40	*	Cancer
6	130**	≥40	*	Heart Disease
7	130**	≥40	*	Viral Infection
8	130**	≥40	*	Viral Infection
9	130**	3*	*	Cancer
10	130**	3*	*	Cancer
11	130**	3*	*	Cancer
12	130**	3*	*	Cancer

3. NON-COORDINATED SYSTEM

Our paper focuses on the publication of data in a non-coordinate system, so understanding the non-coordinate system is essential. In simple terms, a non-coordinated system is simply a system that has no prior coordination between themselves before releasing data.

Due to an increase in decision-based software, collecting specific information has become a necessity. Still, when sharing these data, the organizations must make sure to respect the privacy of individuals. At first, the privacy-preserving methods were only focused on a single instance of data release, but they were not enough when several different organizations published data. Then the research moved on to serial publication [5][13]. In that case, the study revolved around several publications that were still under one organization, so the publishers had prior knowledge about the previous data sets. But this paper focuses on the publication of independent organizations publishing data where some individuals may be common. This is a non-coordinated publication system as one publisher does not know what other publishers may release or if they have any common data sets.

4. PROPOSED METHODS TO MITIGATE COMPOSITION ATTACK

Over the years, many techniques have been proposed to prevent composition attacks. None of the methods have been a hundred percent successful, but some have been able to lessen the risk factor to a certain degree without compromising the data set's utility. In our paper, we have summarized the five most established methods after much consideration. These are discussed below in detail.

Table 2: List of Methods Surveyed

Author	**Method**	**Year**
Ganta et al. [11]	Partition Based Schemes	2008
Baig et. al. [16]	(ρ, α)-Anonymization Generalization	2012
Sattar et. al. [17]	Probabilistic Approach	2014
Li et. al. [18]	Sampling, Perturbation and Generalization	2016
Hasan et. al. [19]	Cell Generalization & Merging Anonymization	2018

The composition attack was first explained and defended with partition based method in [11]. Not much progress had been made after that until 2012 when (ρ, α)-Anonymization Generalization [16] was developed. The generalization technique was later used in [18][19] but with much improved results. The latest study to prevent the composition attack was in 2018.

4.1 Partition Based Schemes [11]

The partition-based scheme deals with the scenario when the publisher is not aware of other anonymized publications. They study the success of insertion attacks, which is an exemplar of a composition attack, empirically. By running an insertion attack on two popular anonymized data sets using partition based schemes, the severity of the attack is measured. It is proven that previous anonymized techniques such as k-anonymity and its recent variants, l-diversity, and t-closeness, are indeed vulnerable to a security breach. The insertion attack relies on two properties of the partition-based anonymization schemes: *1.Exact sensitive value disclosure:* the sensitive value corresponding to each member of the group is published precisely. *2.Locatability:* given any individual's non-sensitive values, one can find the group in which the individual has been put. Both properties are

widespread. The exact sensitive value disclosure is common to feature all the arrangements based on k-anonymity. Locatability is less well known since it rests on the precise choice of partitioning algorithm and the non-sensitive attributes recording. Still, some methods always fulfil locatability like- schemes that recursively partition the data set along the lines of a hierarchy that is later used for generalization or techniques that release the same set of non-sensitive attribute vectors for each group. Taking that into consideration, this research suggests that using simple heuristics one can locate an individual's group with high probability.

In partitioning-based anonymization, the attributes are divided into two classes. These are- sensitive attributes and non-sensitive attributes. Any sensitive attribute is private to the individual and should not be made known to everyone. All other attributes are non- sensitive attributes, which are also dubbed as Quasi-identifiers.

At first, the scheme clusters individuals into groups. Then it generalizes the non-sensitive values so that each group forms an equivalence class relating to the quasi-identifiers depending on different criteria. Here l-diversity and t-closeness are also considered alongside k-anonymity. However, the actual anonymity achieved is less than ideal and is equal to several distinct values in each equivalence class. This is called effective anonymity. Another new term introduced is the vulnerable population (VP) that is the number of individuals for whom the intersection attack leads to a positive drop in effective anonymity.

To measure the extent of damage probable through the intersection attack [11], two possible situations are considered. These are -

Perfect Breach*:* A perfect breach happens when an adversary can deduce the exact sensitive value of an individual. In other words, when the enemy has a confidence level of 100% about the individual's sensitive data, it is a perfect breach. There are three scenarios for anonymizing the two overlapping subsets-

1. Mondrian on both the data subsets.
2. Micro aggregation on both the data subsets
3. Mondrian on the first subset and micro aggregation on the second subset.

(k1,k2) represents the pair of k values used to anonymize the first and the second subset, respectively Table 3(a) and Table 3(b). In the experiments, the same k values for both the subsets was used, meaning k1=k2. In the case of an Adult database, it was found that around 12% of the population is vulnerable to a perfect breach, and for the IPUMS database, this value is around 60%. As the value of k increases, the percentage of the vulnerable population goes down. The reason for that is that as the value of k increases, the partition sizes in each subset increases. This leads to a larger intersection set and thus lesser probability of obtaining an intersection set of size 1.

Table 3(a): Adult census database

Attribute	Domain Size	Class
Age	74	Quasi ID
Work Class	7	Quasi ID
Education	16	Quasi ID
Marital Status	7	Quasi ID
Race	5	Quasi ID
Gender	2	Quasi ID
Native Country	41	Quasi ID
Occupation	14	Sensitive

Table 3(b): IPUMS census database

Attribute	Domain Size	Class
Age	100	Quasi ID
Work Class	5	Quasi ID
Education	10	Quasi ID
Marital Status	6	Quasi ID
Race	7	Quasi ID
Gender	2	Quasi ID
Native Country	113	Quasi ID
Occupation	247	Sensitive

***Partial Breach*:** If the adversary cannot guess with surety the sensitive attribute of an individual, it is called a partial breach. Meaning the adversary infers the result to a fewer number but cannot be 100% sure. Though in most cases, the few values the adversary finds could reveal a lot of information. For example, for a hospital database, by boiling down the sensitive values of the diagnosis to a few values like- Flu, Fever or Cold, it could be concluded that the person is suffering from a viral infection. In this case, the adversary's confidence level is 1/3= 33%. Here, only the first anonymization scenario described earlier is used. Where both the overlapping subsets of the database are anonymized using Mondrian multidimensional technique. It was noted that the severity of the attack increases alarmingly for slight relaxation on the required confidence level. For example, in the case of the IPUMS database, around 95% of the population was vulnerable to a confidence level of 25% for k1 = k2 =5. Still, for the Adult database, more than 60% of the population was affected.

4.2 Probabilistic Approach [17]

The Probabilistic approach, called (d , α)-linkable, works to improve privacy without any coordination between publications. The model makes sure that d confidential values are associated with a quasi-identifying group with the possibility of α. Thus, the goal of this method is to protect a person's privacy when the data records are released by separate organizations when coordination is not possible. The model designed increases the chance that an adversary will have more than one sensitive attribute to link with an individual's quasi identifier after joining disparate k-anonymized data sets. This is achieved by using statistical information in regards to the quasi identifier and private attributes of the underlying populace to simulate a k-anonymized data set published by another organization.

If we study example 1, we will see that the composition attack is successful for the k-anonymized data sets because an adversary can focus only on one sensitive attribute, 'AIDS'. This is used to link with Alice's record. In comparison, the

adversary has all values from the confidential attribute's domain to connect with Alice's record from the pair of differentially private data sets. This property is labelled as *likability*, which enables attackers to narrow down the search of a victim's sensitive values. Therefore, when the adversary knows that Alice went to both hospitals, they can guess that Alice has 'AIDS' with 100% confidence. The level of likability is determined by the number of confidential values shared across multiple k-anonymized data sets in an individual's equivalence class. Suppose *d* represents the value of common confidential values. In that case, the likability of those anonymous data set is *d*, and one anonymous data set is *d*-linkable with another anonymous data set. When the number of same confidential values increases, the risk of an individual's privacy being breached lessens.

The algorithm for this method is called "dLink". Basically how the algorithm works is when an equivalence class fails to satisfy the (d, α)-linkable model [17], it is merged with its closest equivalence class (For a k-anonymized data set, an equivalence class is the set of records in the data set that has identical values of quasi-identifiers). This process keeps repeating until the model is satisfied. The merging of the equivalence class can be done in one of two ways, and these are- first by increasing the dimension of the equivalence classes and re-anonymize the original data set. Or the equivalence class that can not satisfy the privacy criterion can be merged with another class that fails to meet the benchmark as well.

Although the tests show that the dLink algorithm can significantly reduce the likelihood of a composition attack, it does not completely eliminate it. The proposed (d, α)-linkable model requires that a data set be published such that each equivalence class contain *d* confidential values with a certain likelihood, but that is not always possible. Also, this model is based on the assumption that all attributes of a record are independent. When that is not the case most of the time, this model was evaluated only in the composition of two data sets. In contrast, in real life, the composition attack may be executed in a more distributed data set.

4.3 Sampling, Perturbation and Generalization [2]

Here the algorithm used is a combination of sampling, perturbation and generalization [18] to protect data privacy from composition attacks. These same methods have been used in previous methods but never combined. This technique significantly reduces the risk of composition attacks and preserves good data utility, which was not the case before.

In generalization, the quasi-identifiers get distorted a little to create an equivalence class where each individual in the same group cannot be separated from each other. Thus, they reach anonymization. Whereas perturbation randomly adds noises to quasi identifier attributes in a data set to reduces the confidence of an adversary of finding an individual's record based on the quasi identifier values. The greater the level of generalization or perturbation, the better it is for privacy protection, but data the utility lowers considerably, which can render a data set useless.

The basis of this technique is dependent on the adversary's chance of making a match in data sets to a specific individual. If the victim's record is in two different data sets and can be revealed with an intersection of the data sets, it is called a *true match*. However, there is the possibility of two separate individuals in two separate data sets having the same record. Then in case of a composition attack, the attacker will match different individuals in two different data sets, and it is called a *false match.* So, a match may be made by two different individuals, and it is called a random match. Even if the victim is not present in two published data sets, there still remains a chance that two records will match. Such a probabilistic inequality protects the privacy of individuals in the non-coordinate system without compromising the data set too much. The more there are shared sensitive values of separate individuals in the data sets by chance, the better the possibility of privacy is against a composition attack. As long as there is more than one shared sensitive value by chance, the attacker could not be sure that the matched records belong to the victim who they are searching for.

Following the above-mentioned principle, a hybrid method is proposed to combine three simple anonymization techniques: sampling, generalization, and perturbation. First, sampling and perturbation are done randomly to generate an ambiguity that links a record to a victim, and then generalization is done to augment the probability of two distinct records seeming to be the same in two data sets. The scheme is used as a preprocessing of an anonymization method that already exists. The time complexity of this algorithm is linear to the dimension and the size of the data set. This way, the probability of a successful composition attack is greatly reduced. In contrast, the anonymized data set maintains utility much more than the differentially private data set.

So, the algorithm at first increases the probability of random match by generalizing the quasi identifier attribute.

Next, the probability of a true match is reduced by sampling and perturbation. When three methods are combined, a data set can be slightly generalized, lightly perturbed and marginally removed, and its utility can be preserved greatly. Another strong point for the hybrid approach is that it is not easy for a rival to reversely build the original data set from an anonymized data set, seeing as the adversary will have to attempt a mixture of three methods.

Let's compare this technique to previous methods. We will find that this method has lowered privacy risks by sacrificing the data quality slightly in contrast with the dLink algorithm. We can understand that this method is similar or better than previous methods.

4.4 Cell Generalization & Merging Anonymization [19]

The cell generalization method is used to raise the privacy of the published dataset, and the other one, margining anonymization, works for increasing the possibility of false matching during composition attack in many independent data publications. It divides the data vertically and horizontally. In the vertical portion, the connected attributes are grouped into a column, and each column will hold a subset of attributes. In the horizontal portion, they are grouped into an equivalence class. For cell generalization, each QI value and l distinct sensitive values are linked together. This paper's proposed method is used to increase the possibility of false matches by linking the QI values with the l distinct sensitive values. Once a patient's record is comparable in two datasets, some common values remain at the intersection of the anonymized data sets, including QI values and sensitive values. When a

patient's record is not available in the two datasets, there may remain a common record in both anonymized datasets. That is induced by two, unlike patients who have the same QI and sensitive values. This type of matching is called false matching.

The main aim of merging the anonymization technique is to increase the possibility of false matching during a composition attack. By using the principle in the variance privacy context, a common record's appearance is independent of whether or not the common record goes to the same individual. An adversary cannot be sure whether the value in the common record goes to the patient. A cell generalization method is presented for protecting the published datasets from composition attacks. The anonymization algorithm is used to anonymize the dataset to confirm the protection from the composition attack and for increasing the data utility.

For solving the problem, the Anonymization algorithm and Privacy Checking algorithm is used. The anonymization algorithm consists of four steps, and they are the creation of fake tuples, attribute separation, tuple separation, and cell generalization. In the anonymization algorithm, it takes a dataset and generates an anonymized table, satisfies l-diversity. After creating fake tuples, add that to the original microdata maintaining two data structure queues. One of the equivalence classes and another set of anonymized equivalence classes. In each iteration, the anonymization algorithm eliminates an equivalence class from a queue of equivalence classes and breaks the equivalence class into two equivalence classes.

By the Privacy-Check algorithm, privacy is checked. After that, two equivalence classes are added to the queue. If the equivalence class cannot be broken, then the anonymization algorithm sets the equivalence class and lastly, the anonymized table is published. The privacy requirement has been declared in every equivalence class by the privacy checking algorithm. For breaking cross-column associations, column values are permuted in the anonymization. There is an opportunity of creating some incompatible tuples in the procedure. Incompatibility tuple is checked, and If there remain incompatible tuples, there generalize a particular cell value to satisfy k-anonymity. To satisfy the privacy requirement, it confirms the l-diversity of all equivalence classes.

Merging anonymization technique offers smaller privacy risks for the composition attacks, and it contains lower relative query error and lower data loss.

4.5 (ρ, α)-Anonymization Generalization [16]

Here discussion is about a different generalization principle (*ρ, α*)-anonymization & composition-based generalization to protect privacy. This generalization principle (*ρ, α*)-anonymization effectively overcomes the privacy concerns for manifold independent data publishing. The other one is a technique that enables the enforcement of privacy in the presence of an overlapping record. For understanding the problem and solution more easily, it is explained with an example below.

Example 2: Consider the overlapping condition of a patient in Figures 4,5 & 6[16]. A patient can easily visit more than one hospital in their area, and sometimes the patient was referred to any hospital by a doctor. Here at Hospital-1, original data contain identifier attribute (Name, SSN etc.) and quasi identifier attribute (Age, Sex, Zip code etc.). Original data are converted into generalized data so that individual is not identifiable. Here Table 5(a) has a 3-anonymous and 2-diverse version from Table (b). 3-anonymity refers that values in the QIDs have a minimum of 3 identical copies and 2-diversity refers to each of such a set has a minimum of 2 distinct values in the sensitive attribute. We can now see the problem of overlapping three hospitals in figure 5(c) between Hospital-1 and Hospital-2. Hospital-3 anonymized its dataset and then released it to Table 6(b). Consider that an opponent knows David's QID, and t David has also visited both Hospital-1 and Hospital-3. The enemy who has the precise QIDs detail of David and tries to deduce David's disease from Tables 4(b) and 5(c). They can discover that David's tuple must have been generalized in the first QID sets of Tables 4(b) and 6(c), individually. These groups include the two mutually sensitive values. So, the adversary cannot get any exact disease that David has contracted. For Eliza, there are also two applicant diseases. In the QID collection of Eliza, no disease in Table 6(a).

The planned method (ρ, α)-anonymization leads to Table 6(c) at Hospital-3. Now the adversary has a 50%chance to deduction the sensitive value of any overlapping record. Here overlap-Anonymized algorithm is used. For that, firstly, the algorithm samples each tuple of original data with generalization data. Then overlap tuple and non-overlap tuples are found. The computation is done in five phases: sampling, division, balancing, assignment and generalize. After that, all tuples are sorted based on QID's and the process continues until assigning all tuples is done. Consider two types of QID groups: overlap QID groups and the other is non-overlap QID Generalization. The algorithm's complexity mainly depends on the computation of tuples, sorting of tuples, and searching the optimal tuples.

Table 4: Generalization at Hospital-1

Table 4(a): Original data

Identifier Attribute	Quasi-Identifiers		Sensitive Attribute
Bob	15	male	B
Hudson	45	male	H
Robi	40	female	G
David	20	male	B
Khan	25	male	C
Victor	50	male	H

Table 4(b): Generalized data

Group ID	Age	Sex	Disease
1	15-25	male	B, B, C
2	40-50	*	G, H, H

Table 5: Generalization at the Hospital-2

Table 5(a): Original data

Name	Age	Sex	Disease
Eliza	40	female	R
Artur	30	male	M
Paul	20	male	M
Noreen	45	female	S
Mathew	15	male	Q
Panama	35	female	T

Table 5(a): Generalized data

Age	Sex	Disease
15-30	Male	M, Q, M
35-45	Female	R, S, T

Table 6. Generalization at the Hospital-3

Table 6(a): Original Data

Name	Age	Sex	Disease
David	20	male	B
Anthony	35	male	C
Rick	30	male	C
Stewart	30	male	L
George	28	male	B
Smith	38	male	W
Eliza	40	female	R

Table 6(b): Generalized data ρ

Age	Sex	Disease
20-30	male	B, C, C, L
30-40	*	B, W, R, S

Table 6(c): ρ with α-overlap

Age	Sex	Disease
15-35	male	B, C, C, L
28-45	*	B, W, R, S

Here the developed (ρ, α)-overlap anonymization model provides an efficient algorithm for adding anonymized datasets to attain (ρ, α)-overlap. After all, it has been shown that the anonymized data sufficiently keep privacy. The algorithm is adopted to compute diversity as the typical generalization principle since it is generally adopted and offers stronger privacy than *k*-anonymity. They are found nearly 90% overlap tuples whose privacy is not preserved at all.

5. CONCLUSION

Finally, we can say the privacy in the non-coordinated system against composition attack has increased due to constant research and new methods over the year. However, still, it has not given us perfect protection. No method has been able to guarantee high accuracy without data loss. But as technology grows, the need for data privacy grows, not lessens. In the future, the publications will become more diverse, and new threats will arise, so there is still a vast area for improvement. Instead of using a non-coordinated system coordinating the publications seems to be a safer option, but such action's feasibility is yet undetermined.

6. REFERENCES

[1] L. Sweeney, "k-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledge based Systems, 2002, pp. 557-570

[2] Hillol Kargupta, Souptik Datta, etal, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", Proceedings of the Third IEEE International Conference on Data Mining. Nov. 2003

[3] A.Machanavajjhala, J.Gehrke, and D.Kifer, et al, "ℓ-diversity: Privacy beyond k-anonymity", In Proc. of ICDE, Apr.2006

[4] LeFevre K, DeWitt D J, Ramakrishnan R, "Mondrian multidimensional K-anonymity", In Proc of the International Conference on Data Engineering(ICDE'06), Atlanta, GA, USA, April.2006, pp. 25-35

[5] K. Wang and B. C. M. Fung. Anonymizing sequential releases. In KDD, pages 414–423. ACM Press, 2006.

[6] C. Dwork. Differential privacy. In ICALP, pages 1–12. Springer, 2006

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

[8] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-anonymity and l-Diversity", In Proc. of ICDE, 2007, pp. 106-115

[9] 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)

[10] Yingyi Bu, Ada Wai-Chee Fu,et al, "Privacy Preserving Serial Data Publishing By Role Composition", In VLDB, 2008, pp. 845-856.

[11] S.R.Ganta, S.P.Kasiviswanathan, Adam Smith : Composition attacks and auxiliary information in data privacy. In: KDD 2008, pp. 265–273 (2008)

[12] Yan Zhao, Ming Du, Jiajin Le : A survey on privacy preserving approaches in data publishing. In: Proceedings - 2009 1st International Workshop on Database Technology and Applications, DBTA 2009, pp.128-131volume 2, Issue 1 (2009)

[13] 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 (2010)

[14] Benjamin C.M. Fung, Ke Wang, Rui Chen, Philip S. Yu : Privacy-preserving data publishing: a survey of recent developments. In: ACM Comput. Surv. 42 (4) (2010)

[15] M.M.Baig, J.Li, J.Liu, H.Wang : Cloning for privacy protection in multiple independent data publications. In: Proceedings of the 20th ACM International Conference

on Information and Knowledge Management, pp. 885–894, Glasgow (2011)

[16] Muzammil Baig, Jiuyong Li, Jixue Liu, Xiaofeng Ding,HuaWang : Data privacy Against composition attack. In:KDD 2012, pp. 320-334

[17] A. H.M. Sarowar Sattar, Jiuyong Li, Jixue Liu, Raymond Heatherly, Bradley Malin : A probabilistic approach to mitigate composition attacks on privacy in non-coordinated environments. In: Knowledge-Based Systems, pp. 361-372 volume 67 (2014)

[18] Jiuyong Li, Muzammil M. Baig, A.H.M. Sarowar Sattar, Xiaofeng Ding, Jixue Liu, Millist W. Vincent : A hybrid approach to prevent composition attacks for independent data releases. In: Information Sciences, 2016, volume 367-368, pp 324-336 (2016)

[19] A S M Touhidul Hasan, Qingshan Jiang, Hui Chen and Shengrui Wang : A new approach to privacy preserving multiple independent data publishing. In:KDD,2018,pp.1-22 volume 8, Issue 5(2018)