

PRIVACY PRESERVING BIG DATA SOCIAL MEDIA ANALYTICS USING ANONYMIZATION AND RANDOMIZATION TECHNIQUES: A SURVEY

¹Andrew. J, Research Scholar, School of Information Technology, VIT University, Vellore
E-mail: yogeshmj.nie@gmail.com

²Dr. J Karthikeyan, School of Information Technology, VIT University, Vellore
E-mail: karthikeyan.jk@vit.ac.in

³Allan Jophus, Student, Computer Science Technology, Karunya Institute of Technology and Sciences, Coimbatore E-mail: allanjophusnewell@gmail.com

Abstract—

Big data refers to huge amount of structured, semi-structured and unstructured data. Due to the rapid growth of online social network users, stupendous amount of data is being generated every day. As the amount of data increases there is a chance of increase in privacy breach of individuals. Social media data usually contains personally identifiable information (PII), processing the social media data without removing the PII would lead to privacy breach of individuals. PII can be removed or protected in different phases of big data life cycle (e.g., data generation, data storage and data processing). This paper is focusing on the survey of preservation of privacy during data processing. Anonymization is one of the globally accepted mechanisms to protect the personal identifiers. Hence the goal of this paper is to give an overview about the various anonymization techniques to protect individual's privacy. This paper also compares the performance of various anonymization techniques like k-anonymity, l-diversity, t-closeness, differential privacy and randomization. The main contributions of this paper include the detailed classification of a good number of recent research works and the demonstration of the recent trend on the security and privacy protection in big data analytics.

Index Terms— Big data, Security, Privacy, Data anonymization, k-anonymity, l-diversity, t-closeness, differential privacy, randomization based privacy preserving

I. INTRODUCTION

Data sets that are so complex and large that the traditional data processing application is inadequate to deal are referred to as Big Data. Big data tends to refer the use of user

analytics behavior, predictive analysis and many more advanced data analytics methods that extract value from data. Big data characteristics are recently explained by 7 V's they are Volume, Velocity, Variety, Variability, Veracity, Visualization and Value. Web 2.0 allows user to interact online, thus there are lot of online social networking sites are created for the online users. Social networking sites provide a platform for the users to interact, share and post their interests. Social networking sites can be divided into social networking sites, social media networks and location based networks. Social networking sites enable the user to connect with various users online. Social media networks allow the user to share their personal data online. Location based networks represents the social networks which is created for a specific group and for a specific locality. Gathering data from social media websites and analyzing the data to predict the future is called Social Media Analytics. The big data life cycle comprises of data generation, data processing, data storage and visualization. Large numbers of personal information are available in big data. This personal information is called as Personal Identifiable Information (PII). PII individually identifies the user. Thus, it is an important task to protect the PII. Unprotected PII may cause privacy breach because the individuals can be targeted for marketing purpose. Security is another important aspect. Security usually enforced at the data storage phase. Protecting data from unauthorized and illegal access is called as data security. Some of the other big data challenges are analysis, storage, capture, data curation, sharing, transfer, search, visualization, updating, querying and information privacy. This paper focuses on information privacy.

Big data has various privacy issues they are, privacy breaches and embarrassments, data masking could be defeated to reveal personal information, anonymization could become impossible, unethical actions based on interpretations, big data analytics are not 100% accurate, discrimination, few legal protections exist for the involved individuals, big data will probably exist forever, concerns for e-discovery, making patents and copyrights irrelevant.

Researchers have developed and identified various techniques to solve the privacy issues of an online user. The following are some of the privacy preserving techniques (i) Anonymization, (ii) Condensation, (iii) Cryptography, (iv) Distributed data mining, (v) Perturbation, (vi) Privacy preserving data mining, (vii) Randomization based privacy preserving. In this paper we are discussing about various data anonymization

techniques and the randomization based privacy preserving techniques. Sanitizing the information whose intent is privacy protection is referred as data anonymization. The process involves either personally identifiable information or encrypting from the data sets making them anonymous. It reduces the risk of unintended disclosure and avoids making them vulnerable. And Randomization based privacy preserving adds randomness to data sets intending to achieve individual privacy. In data anonymization we focus on the techniques k-anonymity, l-diversity, t-closeness, differential privacy techniques along with the randomization based techniques.

II. ANONYMIZATION

Anonymization is the process of removing or hiding sensitive information from the dataset. Database which contains sensitive information are vulnerable to attack. If the database system is compromised there is a huge risk for privacy breach. Hence the personal information need to be protected. Anonymization techniques are widely used to protect privacy in the data. Actual data contains (i) Quasi attributes (ii) Sensitive attributes (iii) Identifiers and (iv) Non-sensitive attributes. Quasi attributes are attributes which can be related with any other attribute to identify a person specifically. Sensitive attributes are attributes which directly points to an individual. Identifiers are personal identifiers like social security number and other information which uniquely identifies a person. Non-sensitive attributes are general attributes. In this paper we are discussing about the various anonymization techniques to hide the quasi, sensitive and personal identifiers.

A. *k*-anonymity

K-anonymity was first proposed in [1]. K-anonymity targets to protect the quasi identifiers. The k-anonymity technique aims to make each record not comparable from a defined number of records. Only if there is k-1 other records that match those attributes for any record of a given set of attributes, the set of data exhibit k-anonymization. There are two different methods in k-anonymity are suppression and generalization. Suppression means the attributes are replaced with special symbols and generalization means the values are replaced with range value. Adequate privacy protection is not provided by k-anonymity if the sensitive values are in an equivalence class as of [2]. In a trajectory data k-anonymity requires every trajectory to be shared by at least k records and most of the data has to be suppressed in order to achieve k-anonymity according to [3]. K-anonymization is prone to some of the common

attacks such as homogeneity attack, similarity attack, background knowledge attack and probabilistic inference attack[4].

The following table compares the various k-anonymity approaches, dataset used, methods and results along with its performance metrics.

Table 1

S. no	Approaches	Dataset used	Method	Result	Performance metrics
1	Distributed randomization[2]	Diabetic nephropathy clinical data	Generalization	Reduces information leak, increases data availability.	-
2	Ontology generalization hierarchy[3]	Sweeney dataset	Generalization and suppression	Does not reveal sensitive information.	Precision metrics
3	Clustering approaches[4]	Adult dataset	Generalization	Measure of information loss is more when different weights are included.	Discernibility metrics, Generalized loss metrics and Normalized Certainty Penalty
4	Crime reporting[5]	Crime dataset	Generalization	Minimize information loss, maximize classification accuracy	Classification accuracy metrics, Information Loss Metrics
5	Crowdsourcing platforms[6]	Census dataset	Generalization and Suppression	Maximize estimated accuracy	Classification metrics and discernibility metrics

6	Web Services[7]	Private web service selection framework	Generalization and Suppression	Incurs complexity overhead, Scale well for large service composition plans	Assertions metrics
7	Multidimensional suppression[8]	Adult dataset	Generalization and Suppression	Higher predictive performance	Information gained metric
8	Data security in cloud computing[9]	Multiple Parties Dataset	Generalization	Improves efficiency of data, Highly scalable	Entropy metrics
9	Big Data Security[10]	Disaster report	Generalization and Suppression	Secure sensitive information	-
10	Location based services[11]	GeoLife	-	High privacy when cost is same, similar privacy for individuals of different groups.	Entropy metrics

B. l-diversity

l-diversity aims to protect the sensitive attributes of a database. In [12] l-diversity has been proposed to protect the k-anonymized dataset from various attacks like homogeneity attack, background knowledge attack, l-diversity is used to reduce the granularity of the data and preserves the privacy of data sets. Due to information leaked by lack of diversity in the sensitive attribute we opt for the technique l-diversity. All the tuples sharing the same values of their quasi-identifiers should have diverse values for their sensitive attributes. In[13] 2-diverse dataset each set of records are indistinguishable and consists of at least two

diagnoses. The dataset is reduced based on the size of data cubes. The reduction is a trade-off between mining algorithms efficiency and data management. This results in gaining more privacy. An equivalence class will have l -diversity if there are at least l “well-represented” values for the sensitive attribute. A table will have l -diversity if every equivalence class of the table has l -diversity. Frequency l -diversity ensures that data analyzers cannot specify each individual’s sensitive values with a confidence greater than $1/l$ as represented in [14]. Privacy-preserving model has been defined with l -diversity in data anonymization where a data holder can share their data with other data users. l -diversity is the solution for lack of diversity in k -anonymity. l -diversity relates every record to at least l “well represented” sensitive values. In [15] the dataset instances are divided into l groups according to bucketization algorithm and suppressed to achieve good generalization thus it preserves privacy.

Table 2

S.no	Approaches	Dataset used	Algorithms	Result	Performance metrics
1	Increasing efficiency of l -diversity[16]	Quasi-identifiers	OKA algorithm, clustering algorithm	Efficiency has been increased	Clusters satisfy l -diversity
2	Privacy on sensitive information using l -diversity [17]	Medical record, Adult database	(a,d) - diversity	Not effective on attribute disclosure	-
3	Enhancing l – diversity[18]	Adult dataset	(k,l, θ) - diversity	Not effective on sensitive attribute disclosure	Discernibility metrics, distance metrics, Information loss metrics, data utility metrics
4	Improved l – diversity for	Adult dataset	L - incognito	Higher diversity and	Distance metrics

	numerical sensitive attributes [19]			strong resistance to homogeneity attack	
5	L – diversity on micro aggregation[20]	Micro-datasets	l-MDAV and k – partition	Improves on sensitive attribute diversity	Euclidean metrics, information loss metrics
6	L – diversity on micro aggregation[21]	Micro-datasets	L – MDAV	Efficiently satisfies l - diversity constraints	Information loss metrics
7	L – diversity on multifarious inference attack [22]	Location data	MDID	Only 60% of the commercial areas are anonymized	Execution time, Information loss metrics
8	Location - based services on l – diversity[23]	Location Data	Distinct l – diversity, Entropy technique, expand cloak technique	Better than improved interval cloak technique	Query anonymization time, Relative anonymization area
9	Randomized addition on sensitive attributes[14]	Adult dataset	Distinct l - diversity	The QIDs are unaffected	Execution time
10	L – diversity on sensor data[24]	Sensor data	Distinct l – diversity	Changing variable in data does not affect performance	Information metrics

In table 2 the different approaches involved in *l*-diversity, the different types of dataset where the algorithms are implemented and the results and performance metrics are described.

C. *t*-closeness

A novel privacy notion called *t*-closeness is proposed in [25]. *t*-closeness is a refinement of *l*-diversity as it has number of limitations. *l*-diversity is vulnerable to attacks like skewness attack and similarity attack. This makes *l*-diversity not so suitable for privacy preservation. *t*-closeness is a refinement of *l*-diversity to preserve privacy in datasets. A dataset is said to have *t*-closeness if the distance between the sensitive attribute distribution in the dataset and the entire dataset distribution is not more than the threshold *t*. *T*-closeness addresses the privacy problems like identity disclosure and attribute disclosure. It reduces the granularity of the data represented. In [26] an extension of *k*-anonymity, *t*-closeness was prepared which protects data against identity disclosure. *T*-closeness puts on an empirical distribution of the confidential attributes. While *k*-anonymity protects against identity disclosure, as mentioned above, it does not protect in general against attribute disclosure, that is, disclosure of the value of a confidential attribute corresponding to an external identified individual. *T*-closeness provides solution for attribute disclosure vulnerability.

Table 3

S.no	Approaches	Dataset used	Technique	Result	Performance metrics
1	Privacy Protection[27]	Netflix Prize data set	Generalization property	Attribute correlations are lost with columns containing small attributes	Correlation metrics
2	Sensitive Quasi-identifiers[28]	Patient table	Subset property	High quality of data within a realistic period	Utility Metrics
3	Micro data[29]	Microdata set	Subset property	Better performance compared to popular deterministic aggregation algorithms	Entropy metrics

4	Enhancing l-diversity[25]	Population dataset	Ordered distance	No sufficient protection for attribute disclosure	Earth mover distance metrics
5	Enhanced Utility preservation[30]	Patient discharge dataset	Microaggregation	Protects sensitive information and is high speed	Information loss metrics
6	Micro data[31]	Microdata set	Subset property	Prevents sensitive data from disclosure	Information loss metrics
7	Data anonymization[32]	Patient data	Subset property	Resists similarity attacks to a certain extent	Distance metrics
8	Data privacy[33]	Adult dataset	Generalization property	Not sufficient protection under attribute disclosure	Discernibility metric
9	Data Publishing[33]	Adult dataset	Generalization property	Though not vulnerable to attacks, data utility loss occurs	Data utility metrics
10	Visual processing[34]	Visual information	Visual abstraction	The error of subject identification results in violation of the subject's privacy	Accuracy of subject identification was 90%

The above table describes the various approaches of t-closeness with its respective algorithm, datasets, results and performance metrics.

D. Differential privacy

In[35] differential privacy has been proposed to guarantee privacy in statistical databases. Differential privacy helps to discover large number of user's usage pattern without compromising privacy of an individual. Differential privacy adds mathematical noise to the individual's usage pattern. The main aim of differential privacy is to increase the correctness of queries from statistical databases and to decrease the chances of record to be identified. According to[36] differential privacy can be provide privacy only when the output of two databases are differing in at most one record. Differential privacy has been a center of focus in privacy protection research. Privacy protection limits the knowledge gain between the datasets that differ among individuals. Differential privacy criticizes the limitation of k-anonymity as it provides poor attribute disclosure. The table is to describe the various approaches and techniques involved in differential privacy.

Table 4

S.no	Approaches	Dataset used	Technique	Result	Performance metrics
1	Trajectory data in participatory sensing[36]	Microsoft research's T-drive project dataset	Hausdorff distance methodology , Bounded noise generation	Privacy loss in the scheme is at least 69%	Hausdorff distance metrics
2	Privacy in smart meters[37]	REDD dataset	Battery based differential privacy scheme (BDP), Cost friendly differential privacy preserving (CDP)	Cost decreases as weight increases	Mutual information metrics
3	Differential privacy on data publishing[38]	Trajectory, NetTrace, Flu, Unemployment datasets	Series-Indistinguishability	High level of data utility	Mean absolute error index
4	Differential privacy based on cell merging[39]	Two dimensional spatial dataset	Grid partitioning	Cell merging is better than uniform and adaptive grids in	Relative deviation

				reducing the deviation of the query	
5	Based on Data Sanitization[40]	Netflix Dataset	1-dimensional mechanism	Lower bounds for maximal expected error	arbitrary metrics, discrete metrics
6	Differential privacy on probabilistic systems [41]	Probabilistic data	Two-level logic	Achieves Bisimilarity data	Infimum metric
7	Differential privacy on complex polytope [42]	$n \times n$ stochastic matrices, finite valued dataset	Matrix Polytope	Differentially private mechanisms have zero column	-
8	On interactive systems[43]	Statistical dataset	Sound Verification technique, Bisimulation, proof technique	Extra leakage arises accounting for bounded memory constraints	Privacy leakage bound
9	Trajectory data on social media[44]	Trajectory dataset	Noise sampling	The amount of perturbation distortion induced through adding random noise is reduced, Utility of released data are extended	Distance metrics

10	Differential privacy on correlated information[45]	Adult dataset, IDS, NLTCs, Search Logs	Correlated iterative-based mechanism	Compared to global sensitivity, privacy is attained rigorously retaining high utility	Distance metrics
----	--	--	--------------------------------------	---	------------------

E. Randomization-based privacy-preserving

Randomization-based privacy-preserving is method of preserving the individual privacy of the dataset by randomizing the data[46]. This method uses collaborative filtering systems. This achieves privacy by adding randomness to the original data and required levels of privacy can be achieved. The privacy preserving methods are grouped into two broad categories, and then frameworks are chosen according to individual's users expectation. It also helps in selecting appropriate privacy preserving methods.

Table 5

S.no	Approaches	Dataset used	Technique	Result	Performance metrics
1	Content confidentiality based on randomization[47]	Corel dataset	Homomorphic encryption based techniques, Feature index randomization techniques	High efficiency using deterministic distance-preserving randomization	Distance metric
2	Link Perturbation on social media[48]	e-mail network dataset	Neighborhood randomization	Link disclosure are limited	Degree centrality, Betweenness centrality, closeness centrality

3	Randomization for Privacy preserving data mining[2]	Clinical data	Distributed randomization	Reduces information leak and increased data availability	Component analysis
4	Attribute disclosure[49]	Adult dataset	Optimal randomization	Better utility preservation	Query answering accuracy
5	Randomization on Shoulder surfing[50]	Key strokes from qwerty keyboard	Individual key randomization	Increase in time for users to complete the typing task	Usability metric
6	Collaborative filtering schemes[46]	Framework	Gaussian random distribution	Helps to select appropriate privacy methods	Bayesian classifier
7	Sql based randomization[51]	Standard Sql	Dynamic randomization	More resistant to sql injection attacks	JDSF
8	On response techniques in PPDM[52]	Adult dataset	Randomized response techniques	Accurate decisions are obtained	Entropy metrics

III. COMPARISION

In this table, an overview along with the comparison of the various anonymization techniques such as k-anonymity, l-diversity and t-closeness with differential privacy techniques and randomization techniques.

S.no	Approaches	Algorithm Used	Privacy attained
1	K – anonymity for de-identification of health data[53]	OLA algorithm	The Datafly and Samarati algorithms had higher information loss

2	Preserving privacy on data publishing using data mining using k-anonymity[54]	Top-Down algorithm, eIncognito algorithm	The distortion ratio from top down algorithm is 3 times smaller than that of eIncognito
3	k-anonymity on Location privacy[55]	Secret Sharing algorithm	At the lowest accuracy level, a significant number of trips are revealed (69% for the whole trip variant and a communication range of 200m)
4	k-anonymity on crowdsourcing database[56]	Mondrian Algorithm	The FKA approaches preserves some distribution features while Mondrian blindly groups tuples
5	k-anonymity privacy preserving for Spatio-temporal data[57]	Cloaking algorithm	The probabilistic method is 78% effective whereas k-anonymity is only 67%
6	l-diversity for anonymizing sensitive quasi identifiers[28]	Anonymization algorithm, Reconstruction algorithm	Adult dataset containing 15 attributes only had 45,222 records after eliminating records from census dataset. Reconstructed values are not sharper
7	Inverted l-diversity for handheld terminals[58]	Evaluating measured radiation patterns	The usable bandwidth was 13% in measurement
8	L-Branch diversity on MQAM[59]	Maximum ratio combining(MRC) and Selection combining(SC)	Results of MRC containing dissimilar channels is equal to that for MRC with i.i.d channels
9	l-diversity against multifarious inference attacks[22]	Precision algorithm, l-diversification algorithm	Preserves 77% of commercial regions
10	Using l-diversity on sensor data[24]	Genetic algorithm	Changing parameters does not affect performance Changing variability of input affects the output
11	Anonymizing sensitive quasi identifiers using t-closeness[28]	Reconstruction algorithm	The clarity of data after eliminating records are poor.

12	T-closeness for data publishing[33]	Checking Algorithm, Incognito algorithm	The anonymized table satisfies 0.15 closeness and the table satisfies 5-anonymity and (1000, 0.15)-closeness
13	T-closeness on visual processing[34]	Active appearance model	For the image of 640*420 the average frame rate was 15.0 fps
14	Preserving microdata using t-closeness[29]	Microaggregation algorithm	Less information is loss and more utility is preserved
15	Differential privacy based on cell merging[39]	Partition algorithm, Merging algorithm	267850 are extracted from 6442864
16	Differential privacy on data using noise estimation[60]	RoD algorithm	There exist linkage attacks for manufactured synthetic dataset
17	Differential privacy for interactive systems[43]	Sanitization algorithm	Privacy can be leaked
18	Correlated differential privacy [45]	Correlated Iteration and Correlated Update function algorithm	Less utility loss
19	Link perturbation through neighborhood randomization [48]	GenProb algorithm	Tends to reduce the in-degree of a high in-degree node and increase the in-degree of a low in-degree node.
20	Using distributed randomization in medical ethics[2]	Apriori algorithm	Avoids leakage of privacy and information loss
21	Sql randomization[51]	Cryptographic algorithm	More resistant to Sql injection attacks
22	Using Randomized response techniques[52]	ID3 algorithm	When θ is near 0.5 randomization level is higher and true information about the original data is better disguised

IV. CONCLUSION

At present privacy in big data is a challenging research issue. The data generated online is growing in an exponential order. Thus, protecting the privacy information in the data is a main concern. Data anonymization is one of the globally accepted method to protect privacy in the dataset. In this paper, a detailed study of various anonymization techniques is presented. This paper also compares the different anonymization techniques results for different datasets. Applying these data anonymization techniques will provide privacy for data records. Information can be made anonymous so that individuals privacy is protected. This paper also discusses about the other privacy protection mechanisms like differential privacy and randomization privacy preserving techniques. We then presented the comparison among the three different privacy preserving mechanisms. s

REFERENCES

- [1] L. SWEENEY, "k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY," *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.
- [2] Y. Xie, Q. He, D. Zhang, and X. Hu, "Medical ethics privacy protection based on combining distributed randomization with K-anonymity," *Proc. - 2015 8th Int. Congr. Image Signal Process. CISP 2015*, no. Cisp, pp. 1577–1582, 2016.
- [3] M. A. Talouki, M. NematBakhsh, and A. Baraani, "K-anonymity privacy protection using ontology," *2009 14th Int. CSI Comput. Conf.*, pp. 682–685, 2009.
- [4] M. I. Pramanik, R. Y. K. Lau, and W. Zhang, "K-Anonymity through the Enhanced Clustering Method," *Proc. - 13th IEEE Int. Conf. E-bus. Eng. ICEBE 2016 - Incl. 12th Work. Serv. Appl. Integr. Collab. SOAIC 2016*, pp. 85–91, 2017.
- [5] M. J. Burke and A. V. D. M. Kayem, "K-anonymity for privacy preserving crime data publishing in resource constrained environments," *Proc. - 2014 IEEE 28th Int. Conf. Adv. Inf. Netw. Appl. Work. IEEE WAINA 2014*, pp. 833–840, 2014.
- [6] S. Wu, X. Wang, S. Wang, Z. Zhang, and A. K. H. Tung, "K-Anonymity for Crowdsourcing Database," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2207–2221, 2014.
- [7] N. Ammar, Z. Malik, B. Medjahed, and M. Alodib, "K-Anonymity Based Approach for Privacy-Preserving Web Service Selection," *Proc. - 2015 IEEE Int. Conf. Web Serv. ICWS 2015*, no. Section III, pp. 281–288, 2015.
- [8] S. Kisilevich, L. Rokach, Y. Elovici, and B. Shapira, "for K-Anonymity," vol. 22, no. 3, pp. 334–347, 2010.

- [9] S. Kavitha, S. Yamini, and R. P. Vadhana, "An evaluation on big data generalization using k-Anonymity algorithm on cloud," *Proc. 2015 IEEE 9th Int. Conf. Intell. Syst. Control. ISCO 2015*, 2015.
- [10] H. Oguri and N. Sonehara, "A k-anonymity method based on search engine query statistics for disaster impact statements," *Proc. - 9th Int. Conf. Availability, Reliab. Secur. ARES 2014*, pp. 447–454, 2014.
- [11] F. Fei, S. Li, H. Dai, C. Hu, W. Dou, and Q. Ni, "A K-Anonymity Based Schema for Location Privacy Preservation," *IEEE Trans. Sustain. Comput.*, vol. 3782, no. c, pp. 1–1, 2017.
- [12] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-diversity: privacy beyond k-anonymity," in *22nd International Conference on Data Engineering (ICDE'06)*, 2006, pp. 24–24.
- [13] S. Kim, H. Lee, and Y. D. Chung, "Privacy-preserving data cube for electronic medical records: An experimental evaluation," *Int. J. Med. Inform.*, vol. 97, pp. 33–42, 2017.
- [14] Y. Sei and A. Ohsuga, "Randomized Addition of Sensitive Attributes for l-diversity."
- [15] K. Mancuhan and C. Clifton, "Support vector classification with l-diversity," *Comput. Secur.*, 2018.
- [16] B. K. Tripathy, "A fast p-sensitive l-diversity Anonymisation algorithm," pp. 741–744, 2011.
- [17] Q. Wang, "(a , d) -Diversity : Privacy Protection Based on l -Diversity," 2009.
- [18] G. Yang, J. Li, S. Zhang, and L. Yu, "An Enhanced l -Diversity Privacy Preservation," no. 61170060, pp. 1115–1120, 2013.
- [19] J. Han, H. Yu, and J. Yu, "An Improved l -Diversity Model for Numerical Sensitive Attributes," no. 60473055.
- [20] H. Jian-min, "An Improved V-MDAV Algorithm for l -Diversity," pp. 735–741, 2008.
- [21] H. Jianmin, C. Tingting, and Y. Juan, "An l -MDAV Microaggregation Algorithm for Sensitive Attribute l -Diversity," 2008.
- [22] S. Miyakawa, N. Saji, and T. Mori, "Location L-diversity against multifarious inference attacks," in *Proceedings - 2012 IEEE/IPSJ 12th International Symposium on Applications and the Internet, SAINT 2012*, 2012, vol. 1, pp. 1–10.
- [23] F. Liu, C. Florida, K. A. Hua, and C. Florida, "Query l -diversity in Location-Based Services," pp. 436–442, 2009.
- [24] A. Abdrashitov and A. Spivak, "Sensor Data Anonymization Based on Genetic Algorithm Clustering with L-Diversity."
- [25] N. Li, "t -Closeness : Privacy Beyond k -Anonymity and -Diversity," no. 3, pp. 106–115, 2007.

- [26] J. Domingo-ferrer and J. Soria-comas, "Knowledge-Based Systems From t -closeness to differential privacy and vice versa in data anonymization," *Knowledge-Based Syst.*, vol. 74, pp. 151–158, 2015.
- [27] S. Yang, L. I. Lijie, Z. Jianpei, and Y. Jing, "Closeness for Privacy Protection," pp. 1491–1494, 2013.
- [28] Y. Sei, H. Okumura, T. Takenouchi, and A. Ohsuga, "Anonymization of Sensitive Quasi-Identifiers for l -Diversity and t -Closeness," vol. 5971, no. c, pp. 1–14, 2017.
- [29] J. Soria-Comas, J. Domingo-Ferrer, D. Sanchez, and S. Martinez, "T-closeness through microaggregation: Strict privacy with enhanced utility preservation," in *2016 IEEE 32nd International Conference on Data Engineering, ICDE 2016*, 2016, pp. 1464–1465.
- [30] I. Table, "A Robust Privacy Preserving Model for Data Publishing," 2015.
- [31] T. Li, N. Li, S. Member, J. Zhang, and I. Molloy, "Slicing : A New Approach for Privacy Preserving Data Publishing," vol. 24, no. 3, pp. 561–574, 2012.
- [32] J. Domingo-ferrer, D. Rebollo-monedero, and J. Forne, "From t -Closeness-Like Privacy to Postrandomization via Information Theory," vol. 22, no. 11, pp. 1623–1636, 2010.
- [33] N. Li, T. Li, and S. Venkatasubramanian, "Closeness : A New Privacy Measure for Data Publishing," vol. 22, no. 7, pp. 943–956, 2010.
- [34] "Privacy protecting visual processing for secure video surveillance," pp. 1672–1675, 2008.
- [35] C. Dwork, "Differential privacy," *Proc. 33rd Int. Colloq. Autom. Lang. Program.*, pp. 1–12, 2006.
- [36] M. Li, L. Zhu, Z. Zhang, and R. Xu, "Achieving differential privacy of trajectory data publishing in participatory sensing," *Inf. Sci. (Ny)*, vol. 400–401, pp. 1–13, 2017.
- [37] Z. Zhang, Z. Qin, and L. Zhu, "Cost-Friendly Differential Privacy for Smart Meters : Exploiting the Dual Roles of the Noise," vol. 8, no. 2, pp. 619–626, 2017.
- [38] H. Wang and Z. Xu, "Knowledge-Based Systems CTS-DP : Publishing correlated time-series data via differential privacy," *Knowledge-Based Syst.*, vol. 122, pp. 167–179, 2017.
- [39] Q. Li, Y. Li, and G. Zeng, "Differential Privacy Data Publishing Method Based on Cell Merging," 2017.
- [40] N. Holohan, D. J. Leith, and O. Mason, "Differential privacy in metric spaces : Numerical , categorical and functional data under the one roof," *Inf. Sci. (Ny)*, vol. 305, pp. 256–268, 2015.
- [41] J. Yang, Y. Cao, and H. Wang, "Differential privacy in probabilistic systems," *Inf. Comput.*, vol. 254, pp. 84–104, 2017.

- [42] N. Holohan, D. J. Leith, and O. Mason, "Extreme points of the local differential privacy polytope Extreme Points of the Local Differential Privacy Polytope," *Linear Algebra Appl.*, no. August, 2017.
- [43] M. C. Tschantz, D. Kaynar, and A. Datta, "Formal Verification of Differential Privacy for," *Electron. Notes Theor. Comput. Sci.*, vol. 276, pp. 61–79, 2011.
- [44] S. Wang and R. O. Sinnott, "Protecting personal trajectories of social media users through differential privacy," *Comput. Secur.*, vol. 67, pp. 142–163, 2017.
- [45] T. Zhu, P. Xiong, G. Li, S. Member, W. Zhou, and S. Member, "Correlated Differential Privacy : Hiding Information in Non-IID Data Set," vol. 10, no. 2, pp. 229–242, 2015.
- [46] Z. Batmaz and H. Polat, "Randomization-based Privacy-preserving Frameworks for Collaborative Filtering," *Procedia - Procedia Comput. Sci.*, vol. 96, pp. 33–42, 2016.
- [47] W. Lu, A. L. Varna, and M. I. N. Wu, "Confidentiality-Preserving Image Search : A Comparative Study Between Homomorphic Encryption and Distance-Preserving Randomization," vol. 2, pp. 125–141, 2014.
- [48] A. A. Yarifard, "Link Perturbation in Social Network Analysis through Neighborhood Randomization," 2015.
- [49] L. Guo, "On Attribute Disclosure in Randomization based Privacy Preserving Data Publishing," 2010.
- [50] A. Maiti, M. Jadliwala, and C. Weber, "Preventing Shoulder Surfing using Randomized Augmented Reality Keyboards," 2017.
- [51] S. A. Mokhov, J. Li, L. Wang, and A. Sqlrand, "Simple Dynamic Key Management in SQL Randomization," 2009.
- [52] W. Du, Z. Zhan, and C. Science, "Using Randomized Response Techniques for Privacy-Preserving Data Mining," pp. 505–510, 2003.
- [53] D. A. A. Myot *et al.*, "A Globally Optimal k-Anonymity Method for the De-Identification of Health Data," pp. 670–682.
- [54] R. C. Wong, A. W. Fu, and K. Wang, "(«) -Anonymity : An Enhanced -Anonymity Model for Privacy-Preserving Data Publishing," pp. 754–759.
- [55] D. Förster, H. Löhr, and F. Kargl, "Decentralized Enforcement of k -Anonymity for Location Privacy Using Secret Sharing," pp. 279–286, 2015.
- [56] S. Wu, X. Wang, S. Wang, Z. Zhang, and A. K. H. Tung, "K-Anonymity for Crowdsourcing Database," vol. 26, no. 9, pp. 2207–2221, 2014.
- [57] S. B. Avaghade and S. S. Patil, "Privacy Preserving for Spatio-Temporal Data Publishing Ensuring Location Diversity Using K- Anonymity Technique," 2015.
- [58] H. Xiao and S. Ouyang, "Design and Performance Analysis of Compact Planar Inverted-L Diversity Antenna for handheld terminals," no. c, pp. 186–189, 2008.

- [59] J. Lu, T. T. Tjhung, and C. C. Chai, "Error Probability Performance of M -Branch Diversity Reception of MQAM in Rayleigh Fading," vol. 46, no. 2, pp. 179–181, 1998.
- [60] H. Chen *et al.*, "Evaluating the Risk of Data Disclosure Using Noise Estimation for Differential Privacy," pp. 339–347, 2017.

