Role of privacy attacks and utility metrics in crowdsourcing for urban data analysis.

Veena Gadad[1][0000-0002-3396-1719], Dr. Sowmyarani C N[2][0000-0001-9549-1663]

Dr. Ramakanth Kumar P[3][0000-0002-2811-1701], Dr. Dayananda P[4][0000-0001-8445-3469]

[1,2,3]Department of Computer Science and Engineering, R.V College of Engineering, Bengaluru, Karnataka, India.
[4]Department of Information Science and Engineering, JSS Academy of Technical Education, Bengaluru, Karnataka, India.

Abstract. In current era, excessive usage of mobile devices and internet people often participate in the surveys, questionnaires, usability tests, performance measures and quantitative reviews. This process of outsourcing the data collection from the crowd is called mobile crowdsourcing. It involves large group of participating people and allows the researcher or analyst to gather data in real time at relatively lower cost when compared to the traditional methods of data collection. Mobile crowdsourcing has applications in idea generation, urban planning and urban mobility, public participation in problem solving and decision making, collective intelligence, crowd wisdom and human computation. There is a threat to individual's sensitive or personal information when the data is shared. Privacy preservation is a major concern in mobile crowdsourcing as enormous amount of data is being collected from the crowd and used for analytics, forecasting and decision making by extracting useful information. These data contain private or sensitive information related to individual/person who owns it. If the data is used in its original form, it may lead to privacy disclosure as it contains person-specific data. Hence, it is the duty of data curator to anonymize the data, before it is published for public use. The original data should be anonymized in such a way that, it should be very challenging for intruder to obtain sensitive information by means of any privacy attack model.

Keywords: Mobile crowdsourcing, Privacy Preserving Data publishing, Data Anonymization, Privacy attack, Data utility, Privacy breach.

1 Introduction

Urban data analysis is a process of collecting, protecting the data and analyzing the data to improve the city living. Even though the traditional methods of data collection such as surveys conducted through person provide detailed information it is time consuming and cost in- efficient [1]. In recent years crowdsourcing, crowd sensing or mobile crowdsourcing are found to be efficient methods [2] of populating the data, on which

the researchers and analyst can work upon and come up with some decision or create policies.

Mobile crowdsourcing is one of the main strategies to carry out real time urban planning tasks such as municipal monitoring, smart city construction and last mile logistics by coordinating with mobile users. However, the success of such outsourcing depends upon how well the crowd workers response and their commitments. Micro Workers [3], Amazon Mechanical Turk [4], crowd SPRING [5] and Google consumer surveys [6] are some of the crowdsourcing tools, they make the task of data collection simpler for individuals as well as business organizations. The advantage of mobile crowdsourcing lies in converting the time-consuming tasks that is expensive and difficult to complete. The tasks are broken down into more manageable tasks and are outsourced to the crowd across the internet, called as microtasks. Figure 1 shows the process of crowdsourced data collection and management system for urban data analysis.

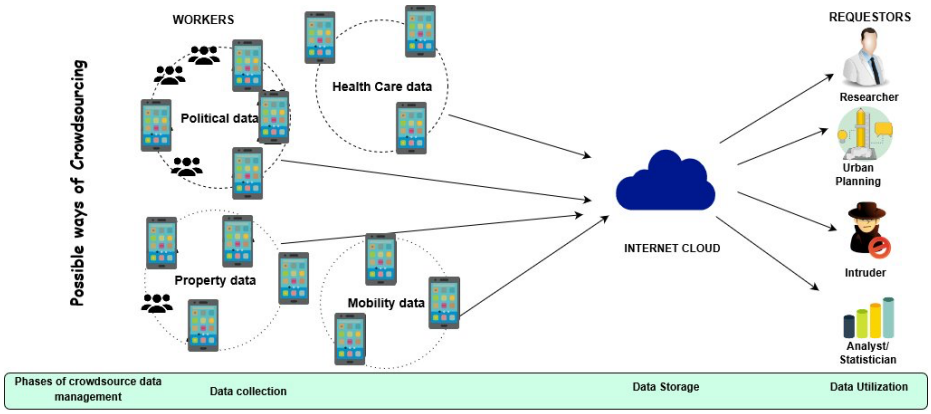

Fig. 1. Crowdsourced data collection and management for urban data analysis.

The crowd maybe human beings or the mobile applications that get involved in conducting various surveys of health care, political data, property information and mobility/transport data. The data thus collected gets stored in a fog/cloud/internet and is utilized for various purposes- planning and decision making, analysis, building models, urban planning and urban mobility, by various users such as researcher, statistician, analyst or an intruder.

Data requestor's job is to publish the task, monitor the task and to collect the answers. The task types may be single choice, multiple choices, fill in the blanks or collection of information. The workers participate in the tasks that are published and sends answers. When series of such tasks are participated by the workers there may be chances of identifying them. For example, in task 1 the survey may be conducted on astrology/horoscope services and hence the date of birth, time and place of birth gets collected. In the next task the survey may be on market survey of a specific product, here the age, workplace and salary of the person is collected. Now when we take the

common people who have participated in both surveys and assign an identifier to them, we clearly get the date of birth, salary and age information of the specific person. Table 1 presents the applications of mobile crowdsourcing and the information collected in such application that may result in disclosure.

Table 1. Applications of mobile crowdsourcing along with possible attributes collected in the process.

Sl.no	Crowdsourcing Application	Description	Attributes collected
1	Smart navigation	Plan route according to weather conditions, accidents, and traffic jams.	Zip code, age, date and time
2	Smart parking	Parking availability with minimum parking charges	Latitude and longitude, data and time.
3	Health monitoring	Status of health	Age, gender, hobbies, work culture
4	Weather condition monitoring	Temperature, Rainfall duration, effects of rainfall.	Date, time, place of data collection.
5	Food recommendation system	Type of food/Drink based on health condition.	Age, gender, hobbies, work culture
6	Horoscopes/Astrology	Astro speak	Age, gender, hobbies, work culture, time and place of birth.
7	Mobility/Traffic	Planning new roads according to the traffic	Latitude and longitude, data and time.

Every crowdsourcing marketplace has its own policies (for example Amazon Turk machine's policies [4]) that prevent the requestor to collect the personally identifiable information (PII) from the workers that disclose the identity of the workers directly. The attributes like age, zip code, salary information that gets collected as a part of survey may not directly identify the individual (such attributes are called quasi identifiers.) but when combined with other data set, there is high probability of individual disclosure. Despite of policies restrictions, it is not possible to prevent the mis use or combination of information. Hence there is a high need for an anonymization technique to protect the individual's disclosure in any form. The basics of any anonymization technique is presented here.

Let A be the original mobile crowdsourced data table, the identifiers (if present) are removed and anonymization methods are applied on Quasi-Identifier's. The anonymized table A` consist of (Quasi-Identifier's and Sensitive Attributes). From the literature, the attributes can be classified and defined as follows:

1. QuasiIdentifiers also called as (QIDs)- Used to identify the individuals but not uniquely for example- person's age, zip code and place of work. This is shown in Table 2.

2. Confidential/sensitive attributes (SA)- Person's sensitive information which needs to be secured and anonymized, for example disease, salary information, political interest etc. as shown in Table 2.

The objective of any anonymization technique is to prevent any third party from identifying an individual

Contribution of the paper- The paper discusses role of privacy attacks and utility metrices by presenting various attacks that may occur in the mobile crowdsourced data and at the same time various available metrices for measuring the information loss that happens due to anonymization. The paper is organized as follows Section 2 presents the existing system. Section 3 provides details of privacy attacks in crowdsourcing. Section 4 discusses the utility metrices for measuring the information losses incurred during the process of anonymization. Section 5 presents conclusion and future work.

2 Existing systems

There is always a tradeoff between data utility and privacy. If we preserve more information without disclosing it in its original form, it leads to less data utility. If the data is disclosed in original form, complete data is utilized which in turn may lead to privacy breach. Data Anonymization uses one or more techniques to make it impossible or difficult to identify a particular individual in the stored data. In order to enhance the utility of the collected data and to preserve the privacy many techniques are available in the literature [7-14]. These techniques use either generalization, suppression or data swapping mechanisms to achieve privacy. For example, consider Table 2, here the zip code and the age are quasi-identifiers, the values are suppressed to prevent further disclosures. K- anonymity [15], is a privacy preserving method that groups similar QID valued attributes into k group, hence Table 3 is 3-anonymous version of table 2.

Table 2. Original table

ID	ZIP code	Age	Disease
1	54677	39	Heart
2	54602	32	Heart Disease
3	54678	37	Heart Disease
4	54905	53	Gastritis

5	54909	62	Heart Disease
6	54906	57	Cancer
7	54605	40	Heart Disease
8	54673	46	Cancer
9	54607	42	Cancer

Table 3. 3-Anonymous version of original table

ID	ZIP	Age	Disease
1	546**	30-40	Heart Disease
2	546**	30-40	Heart Disease
3	546**	30-40	Heart Disease
4	549**	50-70	Gastritis
5	549**	50-70	Heart Disease
6	549**	50-70	Cancer
7	546**	40-50	Heart Disease
8	546**	40-50	Cancer
9	546**	40-50	Cancer

Here k=3, indicates number of records grouped into one class where, QID values are same in all three records leading to 3-anonymity.

Differential Privacy (DP) [16][17] was initially developed for interactive query and response system. The query results are randomized using the distributions like the Laplace, Gaussian or Geometric distributions. The variant of DP is non interactive DP, here the sanitized dataset is released to for public use regardless of type of the requestor. Such non interactive DP measures [18][19] suffer badly with 'curse of dimensionality' which means as the number of dimensions increases with applications of privacy techniques to the individual attributes having high correlations gets weakened [20], [21], this increases the threats as well as reduces the utility. Even worse, the privacy guarantee of DP degrades exponentially when multiple correlated queries are processed. Xuebin Ren et.al [22] uses a Local Differential Privacy (LDP) technique for high dimensional crowd sourced data publication. It is particularly useful in crowdsourced data, where each user contributes the single private data record to an untrusted server. LDP has its own practical applications in collecting user statistics without harming user privacy. For example, RAPPOR [23], it is a Chrome extension. It collects Windows process names and Chrome Homepages from user devices in an LDP manner. Microsoft has deployed a data collection mechanism that is LDP-enabled in Windows Insiders program to collect application usage statistics. Therefore, the users as well as the software companies gets benefitted from the LDP usage because users obviously need of privacy, the appreciation of preserving user privacy may gains positive reputation for companies. Lastly, the intruders may be able to retrieve or even steal the user data, that violates user privacy.

3 Privacy attacks in crowd sourced data

Privacy is major concern in mobile crowdsourcing. The essential entities of mobile crowdsourcing are the data requestors/end users – these entities request the data through the tasks and then utilizes the data provided by the participants. Data workers/ Participants- provides the response by participating in the surveys/collecting the data of their interest. The tasks are the entities that are distributed or shared across the participants. The privacy threat may occur on the participant or an individual may be disclosed on the data provided by the participant. Many privacy attacks have been researched in the literature [49] with respect to different domains. This section provides the overview of the possible privacy attacks in mobile crowdsourced environment.

3.1 Task tracing attack [24][25] occurs on the crowd workers, the crowd workers pulls the tasks from the market place or distribution servers based on their interest. When the tasks are downloaded the worker shares some of the sensitive attributes such as age, location, time, preferences and the type of sensing device. The tasks pulled by the workers include details of traffic information, political surveys, real time weather information etc. By studying the type and preferences of tasks pulled by the worker there is possibility of leakage of the sensitive information of the participants such as age, location, race, organization, location and other related attributes of the participants. However, tracing more than one tasks pulled by the worker and collecting the information about the participant may disclose the sensitive attributes and lead to privacy threat. For example, consider table 4, the crowd worker is an engineering undergraduate student with his original information in university database. Table 5 and Table 6 contain the information of the tasks pulled by the participant and the list of accepted tasks. The tasks may be are related to traffic, weather conditions and recommendation system.

Table 4. Original student information at University

Name	Specialization	Age	Gender	Sensing device	Zip code
Name_1	Civil	26	M	Android	560098
Name_2	Computer Science	23	M	Android	560098
Name_3	Electrical engineering	25	F	Windows	560098
Name_4	Mechanical engineering	27	F	Android	560098
Name_5	Civil Engineering	29	M	Windows	560098

Table 5. Task requirement

Tasks	Age	Gender	Sensing device	Zip code

Task_1	26	M	Android	560098
Task_2	26	M	Android	560098
Task_3	25	F	Windows	560098

Table 6. Accepted Tasks

User_1	Task_1	Task_2
User_2	Task_3	Task_2
User_3	Task_2	Task_3

3.2 Malicious attack. In crowd sourcing, there may be a malicious requestor or malicious worker or a malicious task, such attacks are called as **malicious attacks [24].** It is an intentionally attack projected by the requestor on the participant or vice versa. The requestor creates the malicious tasks and pushes them to the participants and imposes strict limitations to participant attributes or sensing devices. These attacks are also called as *narrow tasking attack [37*] which are malicious and intentionally created to collect specific attributes to violate their privacy. The other variation of malicious attacks are *selective attacks [37]* where in the task may be pushed and assigned to selective group of participants to trace their attributes or to learn about them. If a participant cannot differentiate between genuine and a malicious task he might be under the attack.

3.3 Collusion attack [25] happens when the requestors are conspired. Consider table 7 and table 8 that consist of information collected by the requestor 1 and 2 separately.

Table 7. Partial Data compiled by requestor1

Date	Time	Age	Gender	Zip code
10-9-19	10 am	26	M	560098
10-8-19	11 am	23	M	560097
10-7-19	12 am	25	F	560098
10-6-19	10 am	27	F	560099
10-5-19	10 am	29	M	560098

Table 8. Partial Data compiled by requestor2

Date	Zip code	Gender
10-9-19	560098	M
10-8-19	560097	M
10-7-19	560098	F
10-6-19	560099	F

10-5-19	560098	M

They create the tasks separately and distribute to the workers. The response for each of these tasks do not reveal any identity however when the requesters share the crowd workers information collision attack take place that may lead to disclosure.

3.4 Sybil attacks are common in network domain where in node in the network operates as multiple identities actively at the same time. The same type of attack may occur in crowdsourcing also [26][27]. The requestors may create fake identities to collect more data from the participants. By aggregating or linking the data provided by the participant, the attacker identifies the crowd workers and get access to their sensitive data. It is very difficult for workers to differentiate sybil attackers to the normal requesters. As a measure for such attacks spatial cloaking or perturbation methods are used that perturbs the original location of the participant [38][39][40].

3.5 Background knowledge attack an adversary has some background knowledge of the participant by having access to other data sources such as census, voter's information or medical history. He now acts as the requestor and assigns tasks to the workers. The result obtained may be mapped with prior information to get more knowledge about a specific individual. Consider an example of Kiva micro funds, it is a nonprofit organization that allows people to lend money via internet to low income entrepreneur's and students across 80 countries [41]. The basic objective of the organization is to connect the borrowers and the lenders across the world. The dataset published by the organization is available in Kaggle Dataset's inaugural Data Science for Good challenge [42]. There are 20 columns in the dataset that is publicly available, consisting of sensitive information like purpose of borrowing loan, Number of lenders and Funded amount. Non sensitive attributes such as gender, country and region. Using this data set and with background knowledge of gender, country and region the attacker plots Voronoi diagram and discloses the sensitive information about the individual. As a measure to this attack, privacy preserving using Voronoi Polygon (PP-Voronoi) [43] is used. The participant forms a cloaked region to prevent his actual identity.

3.6 Location based attacks [44]- When the participant participates in the tasking his location information such as home address, working information living habit etc. may be revealed through the sensing device, that the participant doesn't want to disclose. When the task is submitted along with the location information it reveals lot of personal information of the participant [45]. Two examples of such sensing applications are Gigwalk [46] and mCrowd [47]. They provide the marketplace for tasks that can be performed through smartphones such as confirming some products available on the shelves by taking images, verifying the prices of the products, traffic related information, weather conditions etc. **Location homogeneity attack [28] and Location inference attack [29]**- these attacks are based on the location and background knowledge When k-anonymity is used as privacy preserving technique there is possibility of such attacks. The requestor with the background knowledge of some

sensitive information creates the tasks and based on the response provided by the participant he gets more information.

Table 9. Location information

Location	Task type	Age	Zip code
Location 1	Health related	39	54677
Location 2	Health related	32	54602
Location 3	Health related	37	54678
Location 4	Health related	53	54905
Location 5	Health related	62	54909

Consider the table 3 that is publicly available and the information in Table 9 is compiled based on the responses obtained by the participant. The participant pulls the task based on his interests and also updates the location data that is in the form of longitude and latitude. By comparing both the tables sensitive inferences such as type of disease can be drawn.

Table 10 presents the summary of attacks, scenarios when these attacks take place and the counter measures to overcome the attacks that are discussed in this section.

Table 10. Summary of Privacy attacks in mobile crowdsourcing

Sl.no	Privacy Threats	Scenarios	Counter Measures
1.	Task Tracing Attack	Analyzing the tasks pulled by the participant	Anonymization methods and strong policies.
2.	Narrow Tasking	Intentionally projected attack.	Policies and preferences.
3	Selective tasking	Intentionally projected to selective participant.	Policies and preferences.
3.	Collusion Attack	Conspired requestors	Anonymization Methods
4.	Sybil Attack	Intentionally projected	Spatial cloaking, Special transformation Generalization and perturbation
5.	Background knowledge attack	Published data and background knowledge	PP-Voronoi
6.	Location based attacks	Location information	Spatial cloaking

4 Utility metrics for quantifying information loss and privacy.

Mobile crowdsourcing has many advantages like it helps to collect large amount of data samples, speed of data collection, inexpensiveness of data collections and better quality of data collected. The collected data is published for monetary purpose or it is utilized for making decisions and to carry out research. Such a data contains sensitive attributes and quasi identifiers and when published it may result in individual disclosures. Hence there is a need to anonymization techniques that balances between the privacy and utility of the published data. This section discusses the metrics to quantify the information loss that is incurred when carrying out the anonymization.

These metrices can be classified into two categories based on the objectives of anonymization. Privacy level measuring metrices measures the privacy level like how well the technique safeguards the privacy from known privacy breaches. Information loss metrices measures the amount of information loss incurred when the data is processed. In former case higher the value better is the technique, in latter case lower the value better is the technique.

4.1 Privacy Metrices

It is essential to measure the amount of privacy that is preserved by a specific technique. Most of the existing techniques for anonymization are either based on discretization or randomization. In discretization the values of attributes are partitioned into intervals, for example the age attribute with value =8, can be anonymized as interval data [10-20]. In randomization the original value x_i is returned as x_i+r, where r is the random value drawn from some distribution. The first proposed metric to measure privacy level is **confidence level [30].** The metric is used for technique that uses discretization for anonymization, it measures how the original values can be estimated from the anonymized values. If it can be estimated with c% confidence that the original value x lies in the interval [x1,x2], then the width of the interval (x2-x1) defines the amount of privacy at c% confidence level. For example, for the age attribute, if width of interval is 10, the level of privacy for such technique is 10% confidence.
For randomization-based methods, the distribution of random variable is taken into consideration. **Average conditional entropy [31,32]** is the metric based on the concept of information entropy is used to measure the level of privacy.

Let X be original data distribution and Z be noisy distribution, the average conditional privacy of X given Z is $P(X|Z)= 2^{p(X|Z)}$ (1)

The level of privacy may also be measured using **variance** between the original data and the anonymized data [33]. If x is the original variable and y is the distorted variable variance(x-y)/ variance(x) expresses how closely one can estimate original values using the distorted data.

4.2 Information loss Metrices

Privacy preserving techniques reduce the quality of the data, that leads to information loss. Information loss metrices quantify this loss of utility.
Definition: Given two values v and v` where v is original value and v` is treated value, the deviation of v` from v is the information loss.
Loss Metric*[34]* was proposed to determine the amount of loss incurred when generalizations are applied on the categorical data. For example, consider the hierarchy for work class, the ontology of work class tree is shown in Figure 2.

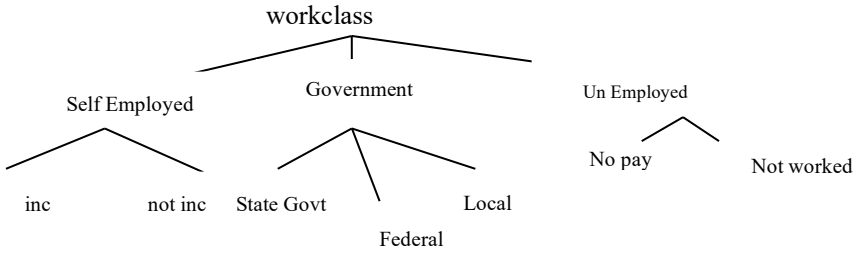

Fig. 2. Ontology of work class tree(T)

Let Q denote number of leaf nodes in the hierarchy tree T and let Q_p denote number of leaf nodes in subtree rooted at P. With generalization, the loss is given by (Qp-1)/(Q-1) For example if the attribute value is State govt and if this attribute is generalized to Government, the amount of loss incurred is computed to be = 2/6. For numeric attributes the loss metric compares the size of generalized domain to the total domain size of attribute.

$$LM_{i,j} = (m-n)/ (max_j - min_j) \quad (2)$$

Here m is the maximum attribute value and 'n' is the generalized or suppressed attribute value in the anonymous table. For the domain of attribute there exist maximum(max) and minimum(min) values. For example, if age is considered as an attribute then the domain range is 1-100. Total loss is summation of loss incurred for individual records.

$$LM(T) = \sum_{i=1}^{|T|} (LM_{i,j}) \quad (3)$$

where, i is the attribute and j are the value of the attribute for an individual record. For Table 3 (anonymous form of table 2), loss for each equivalence class is 0.3, 0.6 and .3 respectively. It can be further computed that total information loss incurred is 13%.
Per record information loss metric [35]- The probabilities of generalized to original are considered to determine the information loss. If a variable B is place of residence that is generalized to B` that could be a state or country then the information loss is given by $\sum_{r\varepsilon G} PRIL(PB, B`, r1, r2)$, where r1 is value taken by B in record r of F and

r2 is value taken by B` in record r of G.For Example: If the place of residence in a record is Madhya Pradesh or Bhopal, it could be generalized to India. P (B=Bhopal | B`= India) < P (B= Madhya Pradesh | B` =India). Since population of India is 1300 million, Madhya Pradesh is 72.6 million and Bhopal is 17.9 million. PRIL scores for P (B=Bhopal | B`= India) = 0.013 and P (B= Madhya Pradesh | B` =India) = 0.055. Lesser this value there is more data loss.

Discernibility Metric (DM) [36] measures number of records that are identical to a given record. The higher the value, the more information that is lost. For example, in the k-anonymity, k −1 record is identical to any given record, therefore the discernibility value will be at least k −1 for any record. More the value k, will increase generalization and suppression, and consequently the discernibility value. For this reason, this metric is considered to be the opposite concept of the k-anonymity. The metric is mathematically represented as

$$\text{DM(m,k)}=\sum_{\forall EQ\ s.t|EQ|\geq k}|EQ|^2+\sum_{\forall EQ\ s.t|EQ|<k}|EQ||T| \quad (4)$$

EQ represents the equivalence class generated by anonymization method m and T represents total number of tuples. The first sum computes penalties for each non suppressed or generalized records and second sum for the suppressed records. In anonymized Table 3, since age attribute of all records are generalized, information loss using discernibility metric is 3*9=27. This indicates that with increase in equivalence class size the information loss also increases.

Many other privacy and information loss metrices are discussed in the literature [48] however the discussed metrics are suitable and easy to evaluate for crowdsourced data. Table 11 shows the summary of privacy loss and information loss metrices.

Table 11. Summary of privacy and information loss metrices in mobile crowdsourced data

Name	Attribute type
Confidence level	Numeric
Average conditional entropy	Numeric
Variance	Numeric
Loss Metric	Categorical and numerical
Per record information loss metric	Numerical
Discernibility metric	Categorical/Numerical

Measuring the information loss is important if the applications that uses the data carry out statistics on the collected information. As discussed previously crowdsourced data finds its applications in many areas, therefore it is essential to anonymize the data as

well as check the amount of loss incurred by application of methods. If there is no utilization of the data then it simply becomes a liability.

5 Conclusion and future work

The essential feature of mobile crowd sourcing is collecting large amount of data efficiently and in a cost-effective manner. Diverse and large work force contribute in performing the task. Hence mobile crowdsourcing finds its applications in problem solving, decision making and wisdom sharing. The privacy of the crowd workers may be at stake and they may be subjected to any of the attacks as discussed in this paper. Therefore, there is a great need of robust privacy preserving technique which is not vulnerable to existing attack. Added to this there is requirement of an efficient privacy preserving technique that protects the privacy and also does not harm the utility of the data. In future, the aim of this study is to explore emerging privacy attacks, evolving existing attacks, to mitigate these attacks by proposing and developing an efficient privacy preserving technique for mobile crowdsourcing.

References

1. Pilloni, Virginia. "How data will transform industrial processes: Crowdsensing, crowdsourcing and big data as pillars of Industry 4.0." *Future Internet* 10.3 (2018): 24.
2. Mansor, M. F., Abdul Halim, H., & Ahmad, N. H. (2018). Exploring crowdsourcing practices and benefits: Validation from small and medium enterprises (SMEs) business owners. In *Proceedings of the 2nd Conference on Technology & Operations Management (2nd CTOM)*. pp. 17-20.
3. Nhatvi Nguyen (2010). *microWorkers*. Retrieved from https://www.microworkers.com.
4. Venky Harinarayan (2001), Amazon turk machine, Retrieved from https://www.mturk.com/
5. Ross (2007). crowdSpring, Retrieved from https://www.crowdspring.com.
6. Google (2012), Google consumer surveys, Retrieved from https://surveys.google.com
7. Xiaoxun Sun; Hua Wang; Lili Sun "Extended K-Anonymity Models Against Attribute Disclosure" Third International Conference on Network and System Security, 2009. NSS '09. pp. 130 – 136.
8. Xiaoxun Sun; Hua Wang; Jiuyong Li; Ross, D. "Achieving P-Sensitive K-Anonymity via Anatomy" ICEBE '09. IEEE International Conference on e-Business Engineering, 2009, pp. 199-– 205.
9. TRUTA T M , VINAY B, "Privacy protection: p-Sensitive k-anonymity property," Proceedings of the 22nd on Data Engineering Workshops, IEEE Computer Society, Washington Dc, 2006.
10. Xiaoxun Sun, Hua Wang, Jiuyong Li, Truta, T.M, Ping Li, "(p+, α) -sensitive k-anonymity: A new enhanced privacy protection model", 2008, pp.59-64.
11. Yingjie Wu; Xiaowen Ruan; Shangbin Liao; Xiaodong Wang; "P-Cover K-anonymity model for Protecting Multiple Sensitive Attributes" The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010

12. Jianmin Han; Huiqun Yu; Juan Yu "An improved l-diversity model for numerical sensitive attributes" Third International Conference on Communications and Networking in China, 2008. ChinaCom 2008. Year: 2008. Pp. 938 – 943.
13. Tripathy, B.K.; Maity, A.; Ranajit, B.; Chowdhuri, D., "A fast p-sensitive l-diversity Anonymisation algorithm," Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE, vol., no., pp.741,744, 22-24 Sept. 2011
14. Qian Wang; Xiangling Shi "(a, d)-Diversity: Privacy Protection Based on l-Diversity WRI World Congress on Software Engineering", 2009. WCSE '09. Volume: 3, pp. 367 – 372.
15. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.
16. Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy: Foundations and Trends in Theoretical Computer Science.
17. Dwork, C., & Lei, J. (2009, May). Differential privacy and robust statistics. In *STOC* (Vol. 9, pp. 371-380).
18. C. Dwork, "Differential privacy: A survey of results," in Proc. TAMC, 2008, pp. 1–19.
19. C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in Proc. TCC, 2006, pp. 265–284.
20. C. Liu, S. Chakraborty, and P. Mittal, "Dependence makes you Vulnerable: Differential privacy under dependent tuples," in Proc. NDSS, 2016, pp. 21–24.
21. J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "PrivBayes: Private data release via Bayesian networks," in Proc. ACM SIGMOD, 2014, pp. 1423–1434.
22. Ren, X., Yu, C. M., Yu, W., Yang, S., Yang, X., McCann, J. A., & Philip, S. Y. (2018). High-Dimensional Crowdsourced Data Publication with Local Differential Privacy. *IEEE Transactions on Information Forensics and Security*, *13*(9), 2151-2166.
23. U. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response," in Proc. ACM CCS, 2014, pp. 1054–1067
24. M. Shin, C. Cornelius, D. Peebles, A. Kapadia, D. Kotz, and N. Triandopoulos, "Anonysense: A system for anonymous opportunistic sensing," Pervasive and Mobile Computing, vol. 7, no. 1, pp. 16–30, 2011
25. L. Pournajaf, D. A. Garcia-Ulloa, L. Xiong, and V. Sunderam, "Participant privacy in mobile crowd sensing task management: A survey of methods and challenges," ACM SIGMOD Record, vol. 44, no. 4, pp. 23–34, 2016.
26. D. Wang, T. Abdelzaher, L. Kaplan, and C. C. Aggarwal, "Recursive fact-finding: A streaming approach to truth estimation in crowdsourcing applications," in Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on. IEEE, 2013, pp. 530–539.
27. K. Zhang, X. Liang, X. Shen, and R. Lu, "Exploiting multimedia services in mobile social networks from security and privacy perspectives," IEEE Communications Magazine, vol. 52, no. 3, pp. 58–65, 2014.
28. Y. Sun, L. Yin, L. Liu, and S. Xin, "Toward inference attacks for k anonymity," Personal and ubiquitous computing, vol. 18, no. 8, pp. 1871–1880, 2014.
29. K. Minami and N. Borisov, "Protecting location privacy against inference attacks," in Proceedings of the 9th annual ACM workshop on Privacy in the electronic society. ACM, 2010, pp. 123–126.
30. R. Agrawal and R. Srikant, ''Privacy-preserving data mining,'' ACM SIGMOD Rec., vol. 29, no. 2, pp. 439–450, 2000.

31. D.Agrawal and C. C. Aggarwal,‘‘On the design and quantification of privacy preserving data mining algorithms,’’ in Proc. 20th ACM SIGMOD SIG ACT-SIGART Symp. Principles Database Syst., 2001, pp. 247–255.
32. Mendes, R., & Vilela, J. P. (2017). Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, *5*, 10562-10582.
33. S. R. M. Oliveira and O. R. Zaıane, ‘‘Privacy preserving clustering by data transformation,’’ J. Inf. Data Manage., vol. 1, no. 1, p. 37, 2010.
34. Iyengar, V.S, “Transforming data to satisfy privacy constraints,”: SIGKDD 2002, pp. 279–288, 2002.
35. Domingo-Ferrer, J., & Mateo-Sanz, J. M. “Resampling for statistical confidentiality in contingency tables,” Computers and Mathematics with Applications, 38(11), 13–32, 1999.
36. R. J. Bayardo and R. Agrawal, ‘‘Data privacy through optimal k-anonymization,’’ in Proc. IEEE 21st Int. Conf. Data Eng. (ICDE), Apr. 2005, pp. 217–228.
37. Shin, M., Cornelius, C., Peebles, D., Kapadia, A., Kotz, D., & Triandopoulos, N. (2011). AnonySense: A system for anonymous opportunistic sensing. Pervasive and Mobile Computing, 7(1), 16-30.
38. Kalnis, P., Ghinita, G., Mouratidis, K., & Papadias, D. (2007). Preventing location-based identity inference in anonymous spatial queries. IEEE transactions on knowledge and data engineering, 19(12), 1719-1733.
39. Wang, S., & Wang, X. S. (2010, May). In-device spatial cloaking for mobile user privacy assisted by the cloud. In 2010 Eleventh International Conference on Mobile Data Management (pp. 381-386). IEEE.
40. Kido, H., Yanagisawa, Y., & Satoh, T. (2005, July). An anonymous communication technique using dummies for location-based services. In ICPS'05. Proceedings. International Conference on Pervasive Services, 2005. (pp. 88-97). IEEE.
41. About Kiva. Retrieved from https://www.kiva.org/about.
42. Elliott Collins (Kiva Impact Team), Retrieved from, https://www.kaggle.com/kiva/data-science-for-good-kiva-crowdfunding#kiva_loans.csv
43. Long, H., Zhang, S., Wang, J., Lin, C. K., & Cheng, J. J. (2017). Privacy preserving method based on Voronoi diagram in mobile crowd computing. International Journal of Distributed Sensor Networks, 13(10), 1550147717739156.
44. Wang, Y., Cai, Z., Chi, Z., Tong, X., & Li, L. (2018). A differentially k-anonymity-based location privacy-preserving for mobile crowdsourcing systems. Procedia Computer Science, 129, 28-34.
45. Wang, X. O., Cheng, W., Mohapatra, P., & Abdelzaher, T. (2013, April). Artsense: Anonymous reputation and trust in participatory sensing. In 2013 Proceedings IEEE INFOCOM (pp. 2517-2525). IEEE.
46. Gigwalk. From http://www.gigwalk.com/
47. Tingxin Yan, Matt Marzilli, Ryan Holmes, Deepak Ganesan, Mark Corner, mCrowd: a platform for mobile crowdsourcing, Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, November 04-06, 2009, Berkeley, California.
48. Wagner, I., & Eckhoff, D. (2018). Technical privacy metrics: a systematic survey. ACM Computing Surveys (CSUR), 51(3), 57.
49. Sowmyarani, C. N., & Dayananda, P. (2017). Analytical Study on Privacy Attack Models in Privacy Preserving Data Publishing. In Security Solutions and Applied Cryptography in Smart Grid Communications (pp. 98-116). IGI Global.