# A NOVEL COMPREHENSIVE CHARACTERIZATION AND QUANTIFICATION TECHINIQUE IN DATA PUBLISHING

Mr. U. Mohan Srinivas[1], Ms. CH. Amrutha[2*]

*1 Associate Professor in MCA Dept., QIS College of Engineering and Technology, Ongole*

*2* Master of Computer Applications, QIS College of Engineering and Technology, Ongole*

**Abstract:** The expanding enthusiasm for gathering and distributing a lot of people's information to open for purposes, for example, medicinal research, showcase examination and affordable measures has made significant security worries about person's delicate data. To manage these worries, numerous Privacy-Preserving Data Publishing (PPDP) procedures have been proposed in this work. In any case, they do not have a legitimate security portrayal and estimation. In this paper, we first present a novel multi-variable protection portrayal and measurement demonstrates. In view of this model, we can dissect the earlier and back antagonistic conviction about property estimations of people. We can likewise investigate the affectability of any identifier in protection portrayal. At that point we demonstrate that security ought not be estimated dependent on one measurement. We exhibit how this could result in security misinterpretation. We propose two unique measurements for evaluation of security spillage, conveyance spillage and entropy spillage. Utilizing these measurements, we dissected the absolute most understood PPDP strategies, for example, k-anonymity, l-diversity and t-closeness. In light of our system and the proposed measurements, we can confirm that all the current PPDP plans have confinements in protection portrayal. Our proposed protection portrayal and estimation system adds to better understanding and assessment of these strategies. Along these lines, this paper gives an establishment to plan and investigation of PPDP plans.

***Keywords:*** *Data Security, Privacy quantification, Data mining, Data privacy.*

## I. INTRODUCTION

These days, datasets are viewed as a profitable wellspring of data for the therapeutic research, advertise investigation and practical measures [3][5]. These datasets can incorporate data about people that contain social, therapeutic, measurable, and client information. Numerous associations, organizations and establishments distribute protection related datasets. While the common dataset gives helpful societal data to analysts, it likewise makes security dangers and protection worries to the people whose information are in the table. To evade conceivable ID of people from records in distributed information, particularly distinguishing data [1], [3], for example, names and government managed savings numbers are commonly expelled from the table.

While the conspicuous individual identifiers are expelled, the semi identifiers [2], for example, postal district, age, and sex may in any case be utilized to extraordinarily recognize a noteworthy part of the populace since the discharged information makes it conceivable to deduce or confine the accessible choices of people than would be conceivable without discharging the table [4]. Truth be told, demonstrated that by connecting this information with the freely accessible side data, for example, data from voter enrollment list for Cambridge Massachusetts [6]-[9], medicinal visits about numerous people could be effectively distinguished . This investigation evaluated that 87% of the number of inhabitants in the United States could be extraordinarily recognized utilizing semi identifiers through side data based assaults [10], including the restorative records of the legislative leader of Massachusetts in the therapeutic information. The spate of security related episodes has prodded a long queue [22], of research in protection thoughts for information distributing and examination, for example, k-secrecy, l-decent variety and t-closeness, to give some examples [11].

A table fulfills k-obscurity if each semi identifier characteristic in the table is indistinct from in any event k □ 1 other semi identifier qualities; such [12]-[14], a table is known as a k-unknown table. While k-secrecy secures personality revelation of people by connecting assaults, it is lacking to avoid characteristic [15], exposure with side data. By consolidating the discharged information with side data, it makes it conceivable to derive the conceivable [16], touchy ascribes relating to a person. When the correspondence between the identifier and the delicate characteristics is uncovered for an individual [17], it might hurt the individual

and the circulation of the whole table. To manage this issue, '- assorted variety was presented in. '- assorted variety necessitates that the delicate traits [18], contain at any rate ' all around spoke to esteems in every equality class. As expressed in '- assorted variety has two noteworthy issues [19]. One, is that it constrains the antagonistic information, while it is conceivable to secure learning of a delicate property from for the most part accessible worldwide conveyance of the quality. Another issue is that all ascribes are thought to be all out, which accept that the enemy either gets all the data [20], or gets nothing for a delicate trait.

In, creators propose a security idea called t-closeness. They initially formalize the possibility of worldwide foundation learning and propose the base model t-closeness. This model requires the dispersion [19] of a touchy property in any comparability class to be near the dissemination of the trait in the general [21], table (i.e., the separation between the two circulations ought to be close to an edge t). This separation was acquainted with measure the data gain between the back conviction and earlier conviction through the Earth Mover Distance (EMD) metric, which is spoken to as the data gain for a particular individual over the whole populace. In any case, the esteem t is a dynamic separation between two conveyances that does not have any instinctive connection with security spillage. Additionally, as we appear in this paper, the separation between two conveyances can't be effectively evaluated by a solitary estimation. T-closeness additionally has numerous restrictions that will be depicted later. The cutting edge PPDP systems will be additionally examined in more subtleties in area.

Research on information protection has simply been centered around security definitions, for example, k-secrecy, l-decent variety, and t-closeness. While these models just consider limiting the measure of security spillage without straightforwardly [18], estimating what the foe may realize, there is an inspiration to discover reliable estimations of how much data is spilled to an enemy by distributing a dataset. In this paper, we start by presenting our novel information distributing system [22]. The proposed system comprises [15], of two stages. To start with, we demonstrate characteristics in a dataset as a multi-variable model. In light of this model, we can re-characterize the earlier and back ill-disposed conviction about trait estimations of people. At that point we portray protection of these people dependent on the security dangers connected with consolidating diverse properties. This model is for sure a progressively exact model to depict protection danger of distributing datasets.

## II LITERATURE SURVEY

### *a)K-anonymity: A model for protecting privacy*

The solution provided in this paper includes a formal protection model named k-anonymity and a set of accompanying policies for deployment. A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. This paper also examines re-identification attacks that can be realized on releases that adhere to k-anonymity unless accompanying policies are respected. The k-anonymity protection model is important because it forms the basis on which the real-world systems known as Datafly, µ-Argus and k-Similar provide guarantees of privacy protection.

(a) Original Table

| | ZIP Code | Age | Disease |
|---|---|---|---|
| 1 | 47677 | 29 | Heart Disease |
| 2 | 47602 | 22 | Heart Disease |
| 3 | 47678 | 27 | Heart Disease |
| 4 | 47905 | 43 | Flu |
| 5 | 47909 | 49 | Heart Disease |
| 6 | 47906 | 47 | Cancer |
| 7 | 47605 | 30 | Heart Disease |
| 8 | 47673 | 36 | Cancer |
| 9 | 47607 | 32 | Cancer |

(b) A 3-anonymous Version

| | ZIP Code | Age | Disease |
|---|---|---|---|
| 1 | 476** | 2* | Heart Disease |
| 2 | 476** | 2* | Heart Disease |
| 3 | 476** | 2* | Heart Disease |
| 4 | 4790* | $\geq 40$ | Flu |
| 5 | 4790* | $\geq 40$ | Heart Disease |
| 6 | 4790* | $\geq 40$ | Cancer |
| 7 | 476** | 3* | Heart Disease |
| 8 | 476** | 3* | Cancer |
| 9 | 476** | 3* | Cancer |

### *b)Robust de-anonymization of large sparse datasets*

Our techniques are robust to perturbation in the data and tolerate some mistakes in the adversary's background knowledge. We apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. We demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, we successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

### *c)T-Closeness: Privacy Beyond k-Anonymity and l-Diversity*

The notion of l-diversity has been proposed to address this; l-diversity requires that each equivalence class has at least l well-represented values for each sensitive attribute. In this paper we show that l-diversity has a number of limitations. In particular, it is neither necessary nor sufficient to prevent

attribute disclosure. We propose a novel privacy notion called t-closeness, which requires that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table (i.e., the distance between the two distributions should be no more than a threshold t). We choose to use the earth mover distance measure for our t-closeness requirement. We discuss the rationale for t-closeness and illustrate its advantages through examples and experiments.

***t-Closeness: Privacy Beyond k-Anonymity and l-Diversity***

The k-anonymity privacy requirement for publishing microdata requires that each equivalence class (i.e., a set of records that are indistinguishable from each other with respect to certain "identifying" attributes) contains at least k records. Recently, several authors have recognized that k-anonymity cannot prevent attribute disclosure. The notion of l-diversity has been proposed to address this; l-diversity requires that each equivalence class has at least l well-represented values for each sensitive attribute. In this paper we show that l-diversity has a number of limitations. In particular, it is neither necessary nor sufficient to prevent attribute disclosure. We propose a novel privacy notion called t-closeness, which requires that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table (i.e., the distance between the two distributions should be no more than a threshold t). We choose to use the earth mover distance measure for our t-closeness requirement. We discuss the rationale for t-closeness and illustrate its advantages through examples and experiments.

## III EXISTING SYSTEM

The spate of privacy related incidents has spurred a long line of research in privacy notions for data publishing and analysis, such as k-anonymity, l-diversity and t-closeness, to name a few . A table satisfies k-anonymity if each quasi-identifier attribute in the table is indistinguishable from at least $k - 1$ other quasi-identifier attributes; such a table is called a k-anonymous table. To deal with this issue, `-diversity was introduced in [4]. `-diversity requires that the sensitive attributes contain at least ` well represented values in each equivalence class. As stated in [5], `-diversity has two major problems. One is that it limits the adversarial knowledge, while it is possible to acquire knowledge of a sensitive attribute from generally available global distribution of the attribute.

***i) Disadvantages***

While k-anonymity protects identity disclosure of individuals by linking attacks, it is insufficient to prevent attribute disclosure with side information. By combining the released data with side information, it makes it possible to infer the possible sensitive attributes corresponding to an individual. Once the correspondence between the identifier and the sensitive attributes is revealed for an individual, it may harm the individual and the distribution of the entire table.

Another problem is that all attributes are assumed to be categorical, which assumes that the adversary either gets all the information or gets nothing for a sensitive attribute.

(a) Original Dataset

| | Zip Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 47677 | 29 | 3K | gastric ulcer |
| 2 | 47602 | 22 | 4K | gastritis |
| 3 | 47678 | 27 | 5K | stomach cancer |
| 4 | 47905 | 43 | 6K | gastritis |
| 5 | 47909 | 52 | 11K | flu |
| 6 | 47906 | 47 | 8K | bronchitis |
| 7 | 47605 | 30 | 7K | bronchitis |
| 8 | 47673 | 36 | 9K | pneumonia |
| 9 | 47607 | 32 | 10K | stomach cancer |

(b) A 3-diverse Version of Salary/Disease

| | Zip Code | Age | Salary | Disease |
|---|---|---|---|---|
| 1 | 476** | 2* | 3K | gastric ulcer |
| 2 | 476** | 2* | 4K | gastritis |
| 3 | 476** | 2* | 5K | stomach cancer |
| 4 | 4790* | $\geq 40$ | 6K | gastritis |
| 5 | 4790* | $\geq 40$ | 11K | flu |
| 6 | 4790* | $\geq 40$ | 8K | bronchitis |
| 7 | 476** | 3* | 7K | bronchitis |
| 8 | 476** | 3* | 9K | pneumonia |
| 9 | 476** | 3* | 10K | stomach cancer |

## IV PROPOSED SYSTEM

All previous approaches to characterize and quantify privacy have only investigated the privacy risk of publishing a sensitive attribute by focusing only on the change of belief of an adversary about the probability distribution of this attribute. However, we believe that any attribute by itself is not sensitive. The sensitivity of an attribute comes from combining it with other attributes. For example, cancer in a medical records dataset, high or low salaries in an employees dataset, are not sensitive unless they are linked to a certain geographical area, age-range or race. To obtain a meaningful definition of data privacy, it is necessary to characterize and quantify the knowledge about sensitive attributes that the adversary gains from observing the published dataset taking into consideration the combinational relation of different attributes.In our approach to characterize privacy, we employ a multi-dimensional scheme of privacy risk analysis attached with combining different attributes. Thus, we introduce the following combinational characterization of privacy.

***Advantages:***

- Privacy-preserving, the publishing technique strictly prohibits any privacy leakage in the published data.

- We focus on instances where different PPDP techniques assume to achieve an intended privacy level.
- We investigate the effectiveness of different PPDP techniques based on our privacy metrics. Simulation results give us a more insightful understanding of privacy leakage.

**Proposed Entropy**

$$\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\geq\mathscr{L}_E(\mathscr{A},\mathscr{C}).$$

*Proof:* We split the proof into four cases.

Case 1: $\sum_{i=1}^m a_i\log_2 a_i \leq \sum_{i=1}^m b_i\log_2 b_i$ and $\sum_{i=1}^m b_i\log_2 b_i \leq \sum_{i=1}^m c_i\log_2 c_i$. Then we have $\sum_{i=1}^m a_i\log_2 a_i \leq \sum_{i=1}^m c_i\log_2 c_i$, and

$$\begin{aligned}&\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\\&=-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m b_i\log_2 b_i-\sum_{i=1}^m b_i\log_2 b_i+\sum_{i=1}^m c_i\log_2 c_i\\&=\left|-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m c_i\log_2 c_i\right|=\mathscr{L}_E(\mathscr{A},\mathscr{C}).\end{aligned}$$

Case 2: $\sum_{i=1}^m a_i\log_2 a_i \geq \sum_{i=1}^m b_i\log_2 b_i$ and $\sum_{i=1}^m b_i\log_2 b_i \geq \sum_{i=1}^m c_i\log_2 c_i$. Then we have $\sum_{i=1}^m a_i\log_2 a_i \geq \sum_{i=1}^m c_i\log_2 c_i$, and

$$\begin{aligned}&\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\\&=\sum_{i=1}^m a_i\log_2 a_i-\sum_{i=1}^m b_i\log_2 b_i+\sum_{i=1}^m b_i\log_2 b_i-\sum_{i=1}^m c_i\log_2 c_i\\&=\left|-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m c_i\log_2 c_i\right|=\mathscr{L}_E(\mathscr{A},\mathscr{C}).\end{aligned}$$

Case 3: $\sum_{i=1}^m a_i\log_2 a_i \leq \sum_{i=1}^m b_i\log_2 b_i$ and $\sum_{i=1}^m b_i\log_2 b_i \geq \sum_{i=1}^m c_i\log_2 c_i$. Then we have $\prod_{i=1}^m a_i^{a_i} \leq \prod_{i=1}^m b_i^{b_i}$, and $\prod_{i=1}^m b_i^{b_i} \geq \prod_{i-1}^m c_i^{c_i}$, and

$$\begin{aligned}&\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\\&=-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m b_i\log_2 b_i+\sum_{i=1}^m b_i\log_2 b_i-\sum_{i=1}^m c_i\log_2 c_i\\&=\log\left(\prod_{i=1}^m \frac{b_i^{b_i}}{a_i^{a_i}}\cdot\frac{b_i^{b_i}}{c_i^{c_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{b_i^{b_i}}{a_i^{a_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{c_i^{c_i}}{a_i^{a_i}}\right)\\&=-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m c_i\log_2 c_i.\end{aligned}$$

Case 4: $\sum_{i=1}^m a_i\log_2 a_i \geq \sum_{i=1}^m b_i\log_2 b_i$ and $\sum_{i=1}^m b_i\log_2 b_i \leq \sum_{i=1}^m c_i\log_2 c_i$. Then we have $\prod_{i=1}^m a_i^{a_i} \geq \prod_{i=1}^m b_i^{b_i}$, and $\prod_{i=1}^m b_i^{b_i} \leq \prod_{i-1}^m c_i^{c_i}$, and

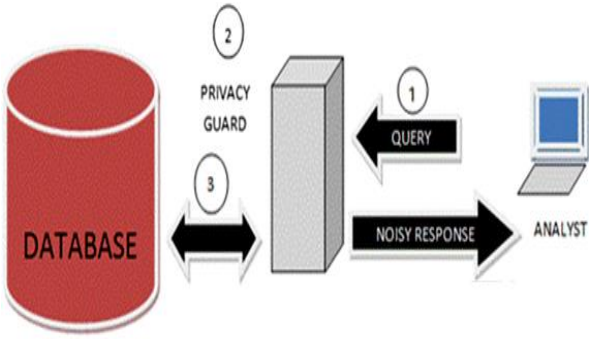
$$\begin{aligned}&\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\\&=\sum_{i=1}^m a_i\log_2 a_i-\sum_{i=1}^m b_i\log_2 b_i-\sum_{i=1}^m b_i\log_2 b_i+\sum_{i=1}^m c_i\log_2 c_i\\&=\log\left(\prod_{i=1}^m \frac{a_i^{a_i}}{b_i^{b_i}}\cdot\frac{c_i^{c_i}}{b_i^{b_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{a_i^{a_i}}{b_i^{b_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{a_i^{a_i}}{c_i^{c_i}}\right).\\&=\sum_{i=1}^m a_i\log_2 a_i-\sum_{i=1}^m c_i\log_2 c_i.\end{aligned}$$

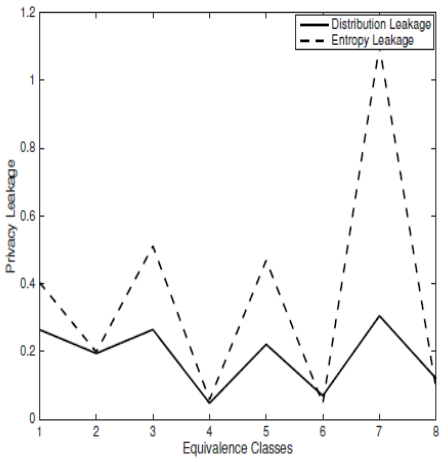
Similarly, we also have

$$\begin{aligned}&\mathscr{L}_E(\mathscr{A},\mathscr{B})+\mathscr{L}_E(\mathscr{B},\mathscr{C})\\&=\log\left(\prod_{i=1}^m \frac{a_i^{a_i}}{b_i^{b_i}}\cdot\frac{c_i^{c_i}}{b_i^{b_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{c_i^{c_i}}{b_i^{b_i}}\right)\geq\log\left(\prod_{i=1}^m \frac{c_i^{c_i}}{a_i^{a_i}}\right)\\&\geq-\sum_{i=1}^m a_i\log_2 a_i+\sum_{i=1}^m c_i\log_2 c_i.\end{aligned}$$

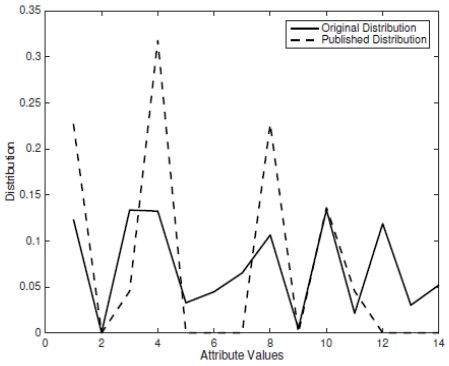
## V METHODOLOGY

The system architecture of our proposed system is given by




(a) Distribution and Entropy Leakage




(b) Original Vs. Published Distribution of a Specific Class

Evaluation of table satisfying 0.5 closeness, 6 diversity, K≥6 anonymity and n=2

## VI RESULT

Privacy-Preserving Data Publishing (PPDP) techniques have been proposed in literature. However, they lack a proper privacy characterization and measurement. We first present a novel multi-variable privacy characterization and quantification model. Based on this model, we are able to analyze the prior and posterior adversarial belief about attribute values of individuals. We can also analyze the sensitivity of any identifier in privacy characterization. Then we show that privacy should not be measured based on one metric. We demonstrate how this could result in privacy misjudgment. We propose two different metrics for quantification of privacy leakage, distribution leakage and entropy leakage. Using these metrics, we analyzed some of the most well-known PPDP techniques such as k-anonymity, l -diversity and t closeness. Based on our framework and the proposed metrics, we can determine that all the existing PPDP schemes have limitations in privacy characterization

## VII CONCLUSION

In this paper, we introduced comprehensive characterization and novel quantification methods of privacy to deal with the problem of privacy quantification in privacy preserving data publishing. In order to consider the privacy loss of combined attributes, we presented data publishing as a multi-relational model. We re-defined the prior and posterior beliefs of the adversary. The proposed model and adversarial beliefs contribute to a more precise privacy characterization and quantification. Supported by insightful examples, we then showed that privacy could not be quantified based on a single metric. We proposed two different privacy leakage metrics. Based on these metrics, the privacy leakage of any given PPDP technique could be evaluated.Our experiments demonstrate how we could gain a better judgment of existing techniques and help analyze their effectiveness in reaching privacy.

Our work opens doors to a wide range of research problems and questions including whether two metrics are sufficient to evaluate privacy or there exist other independent metrics that could help achieve better privacy quantification. Another open problem is the optimization of the original data generalization as to achieve maximum privacy based on our proposed metrics. Typically, we believe that equivalence classes should be designed in such a way that keeps both the entropy leakage and the distribution leakage below a certain pre-determined level. This motivates us to think of a typical publishing scenario. We also leave as an open problem for further research, optimization of the chosen set of quasi-identifiers with an objective of minimizing distribution and entropy leakages within the published table or specific classes of higher privacy concerns.

## VIII REFERENCES

[1] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, pp. 557–570, 2002.

[2] L. Sweeney, "Uniqueness of simple demographics in the U.S. population," 2000.

[3] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in Security & Privacy, pp. 111–125, 2008.

[4] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam, "`-diversity: Privacy beyond k-anonymity," ACM Trans. Knowl. Discov. Data, vol. 1, Mar. 2007.

[5] N. Li, T. Li, and S. Venkatasubramanian, "t-closeness: Privacy beyond k-anonymity and l-diversity," in ICDE, pp. 106–115, 2007.

[6] N. Li,W. Qardaji, D. S. Purdue, Y.Wu, andW. Yang, "Membership privacy: A unifying framework for privacy definitions," in CCS, (Berlin, Germany), 2013.

[7] I. Wagner and D. Eckhoff, "Technical privacy metrics: a systematic survey," CoRR, vol. abs/1512.00327, 2015.

[8] J. Zhang, G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao, "Privbayes: Private data release via bayesian networks," in Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, SIGMOD '14, (New York, NY, USA), pp. 1423–1434, ACM, 2014.

[9] M. G¨ otz, S. Nath, and J. Gehrke, "Maskit: Privately releasing user context streams for personalized mobile applications," in Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data, SIGMOD '12, (New York, NY, USA), pp. 289–300, ACM, 2012.

[10] Y. Rubner, C. Tomasi, , L. J., and Guibas, "The earth mover's distance as a metric for image retrieval," International Journal of Computer Vision, vol. 40, no. 2, pp. 99–121, 2000.

[11] D. Rebollo-Monedero, J. Forne, and J. Domingo-Ferrer, "From tcloseness-like privacy to postrandomization via information theory," IEEE Trans. on Knowl. and Data Eng., vol. 22, pp. 1623–1636, Nov. 2010.

[12] L. Sankar, S. R. Rajagopalan, and H. V. Poor, "Utility-privacy tradeoffs in databases: An information-theoretic approach," Trans. Info. For. Sec., vol. 8, pp. 838–852, June 2013.

[13] C. Dwork., "Differential privacy," ICALP, 2006.

[14] K. Lefevre, D. J. Dewitt, and R. Ramakrishnan, "Incognito: Efficient full-domain k-anonymity," In Proceedings of ACM SIGMOD, pp. 49–60, 2005.

[15] P. Samarati, "Protecting respondents identities in microdata release," IEEE Transaction Knowledge Data Engineering, vol. 13, no. 6, pp. 1010–1027, 2001.

[16] R. J. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymization," In Proceedings of the 21st IEEE International Conference on Data Engineering (ICDE), pp. 217–228, 2005.

[17] B. C. M. Fung, K.Wang, and P. S. Yu, "Top-down specialization for information and privacy preservation," In Proceedings of the 21st IEEE International Conference on Data Engineering (ICDE), pp. 205–216, 2005.

[18] B. C. M. Fung, K. Wang, and P. S. Yu, "Anonymizing classification data for privacy preservation," IEEE Trans. Knowl. Data Engin., vol. 19, no. 5, pp. 711–725, 2007.

[19] V. S. Iyengar, "Transforming data to satisfy privacy constraints," In Proceedings of the 8th ACM SIGKDD, pp. 279–288, 2002.

[20] X. Xiao and Y. Tao, "Personalized privacy preservation," Proc. ACM SIGMOD, pp. 229–240, 2006.

[21] N. Adam and J. Worthmann, "Security-control methods for statistical databases: A comparative study.," ACM Computing Surveys, 1989.

[22] R. Agrawal and R. Srikant, "Privacy-preserving data mining," SIGMOD, 2000.

**Authors Profile**

Mr. **U. Mohan Srinivas** is currently working as an Associate Professor in Department of Master of Computer Applications in QIS College of Engineering & Technology with the Qualification M.Tech (Ph. D).

Ms. **Ch. Amrutha** pursuing MCA 3rd year in Qis College and Engineering and Technology in Department of Master of Computer Applications, Ongole.