ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 5, Issue 12, December 2017

Development of Sensitivity Classification Approach for Personalized Privacy Preservation in Data Publishing (PPPDP)

D.P. Manoj Kumar, Dr. Y.P Gowramma,

Assistant Professor, Department of Computer Science, KIT, Tiptur, India

Professor, Department of Computer Science, Kit, Tiptur, India

ABSTRACT: Privacy preservation in data management and publishing has grown to be a vital research area in the era of big data. Efficiently protecting individual privacy in data publishing is especially critical due to variation in personal preference and sensitivity. The consequences of private data getting published are causing psychological issues and disturbances in individual's personal life. This has triggered the requirement to develop various approaches for privacy preservation in data publishing. Game theory is one of the approaches adopted for privacy preservation in data publishing.The study of situation involving computing interest, modeled in terms of the strategies, probabilities, actions, gains, and losses of opposing players in a game is game theory. A comparison of PPDP with PPDM has been done to explore the utility. Most current studies only manage to achieve personalized privacy preserving in a statistical sense, though some researchers have investigated the issue of data anonymizationusing K anonymity algorithms for personalized privacy protection. K-anonymization techniques have been the focus of intense research in the last few years. In order to ensure anonymization of data while at the same time minimizing the information loss resulting from data modifications At present, personalized privacy preserving in data publishing (PPPDP) is still in the premature stage of development. To form a basis of development, a classification method has been formulated to collect sensitive and non-sensitive data separately, where sensitive data has to be strongly protected. A survey was conducted on the sensitivity of data based on individual's social,economic, personal and psychological factors. A conceptual approach has been designed to achieve personalized privacy preservation in data publishing (PPPDP) based on the classification of sensitivity of individual. Based on the sensitivity classification, Game theory has been proposed for achieving personalized privacy preservation applied in the field of data publishing in the financial and banking sector. This approach can be extended to other sectors of data publishing like social media networks, matrimony data, reviews etc. as future research avenue.

KEYWORDS: - PPDM, PPDP, PPPDP, Senstivity, Game Theory, K-Anonymity, Nash Equilibrium, KDD process

I. INTRODUCTION

Privacy protection becomes an important issue when one wants to make use of data that involves individuals' sensitive information. Research on protecting the privacy of individuals and the confidentiality of data has become one of the major criteria of data security. With the rapid development of information technology and the wide application of networks, large-scale ofdigital information is stored and published. Knowledge discovery and data mining applications in information retrieval are playing an active role and has greatly contributed to the variousdepartments right from data mining of useful information needs. In addition to the benefits of the digital information, many problems regarding the privacy is envisaged. The disclosure of sensitive information has become prominent nowadays, and privacy preservation has become a research hotspot in the field of data security. The association among the illegal records in public security system, the customer's credit card transactions, telecommunications users' personal information, housing information, and soon are some data publishing entities. It is of great significance for government and business organization not to destroy the citizens' personal privacy. A reasonable and effective method of protection, which can protect the user's privacy and keep the data available at the same time, is the trend of developments in information security. (Yn Pen et. al. (2012).The way organizations, such as hospital, government agency, or insurance company, banks can release data to the public without violating the confidentiality of personal information is a challenge of the present day.

It is often necessary to publish personal information for statistical and research purposes. (Xia oui Xiao et. al) The process of privacy preservation in data publishing works in a certain pattern. Data owner offers the data to data publisher, which would have their individual data containing sensitive information. The data publisher treats the collected data uniformly to meet the privacy model. Finally data publisher publishes the data satisfying privacy requirement to the data receiver for statistical analysis. Although the data owner has authorized the data publisher to deal with his individual data for protecting privacy, the uniform treatment of data may not meet privacy requirement of each individual. This leads to problem of personalized privacy preservation (Jinling Song et. al.). This calls for a privacy preservation technique to reduce the possibility of identifying sensitive information about individuals, which is also called as disclosure control problem (Raymond Chi-Wing Wong et. al.).

Benjamin C. M. Fung et. al. (2010) explained that a task of the utmost importance is to develop methods and tools for publishing data in a more hostile environment, so that the published data remains practically useful while individual privacy is preserved. This procedure is called privacy-preservation in data publishing (PPDP).Benjamin C. M. Fung et. al. (2010) accentuated that privacy-preserving technology solves only one side of the quandary. It is equipollent paramount to identify and overcome the non -technical difficulties faced by decision makers when they deploy a privacy-preserving technology. Their typical concerns include the degradation of data/accommodation quality, loss of valuable information, incremented costs, and incremented intricacy. Authors believed that cross-disciplinary research is required to abstract these obstacles, and urge computer scientists in the privacy auspice field to conduct cross-disciplinary research with convivial scientists in sociology, psychology, and public policy studies. Having a better understanding of the privacy quandary from divergent perspectives can work towards personalized privacy preservation technique, which is an identified research gap.LeiXu et.al (2014) has explained in his work on PPDP and PPDM, that it provides method to explore the utility of data while preserving privacy. However, most current studies only manage to achieve privacy preserving in a statistical sense. Considering that the definition of privacy is essentiallypersonalized, developing methods that can support personalizedprivacy preserving is an important direction for thestudy of PPDP and PPDM. Someresearchers have already investigated the issue of personalized anonymization. K anonymisation is one of the most commonly used algorithms for personalised privacy protection. However, the majority of *K*-anonymity algorithms are based on static data sets, and in the real world, data is constantly changing, including changes in forms of data, attributechanges, adding new data, and deleting the old data. Besides, the data between data sets are likely to beinterrelated, how to achieve privacy preservation in a much more complex environment with dynamic data,still need further study. (Yun Pan et. al.(2012). The author has narrated that it is unrealistic to expect every data provider to define his privacy preference in standard pattern, as personalization has become need of the current data driven application. Developing practical personalizedanonymization methods is in urgent need. Further research is required on how to formulate personalized privacy preference in a more flexible way and how to obtain such preference with less effort. Amardeep Singh et.al (2014) Anonymization techniques have been developed for one time released network data.

Xiaokui Xiao et.al. (2006) proposed the concept of personalized anonymity, and developed a new generalization framework that takes into account customized privacy requirements. This technique successfully prevented privacy intrusion in scenarios where the existing approaches fail, and results in generalized tables that permit accurate aggregate analysis. Personalized privacy preservation is still an area of exploration.

AshwinMachanavajjhala et.al(2009).Privacy in data publishing has received much attention recently. A N K Zaman (2014). Concept of this paper the idea of privacy preserving data publishing is discussed for data classification purpose.SlawomirGoryczkaet.al.proposed the concept of considers the collaborative data publishing problem for anonymizing horizontally partitioned data at multiple data providers.

From the above reviews, it is seen that very less work has been done on achieving Personalized Privacy Preservation in Data Publishing (PPPDP). More researches can be seen in respect of privacy security in data mining than that of data publishing.

By taking into account the above facts, in this paper, a conceptual research has been developed with regard to personalized privacy preservation in data publishing.

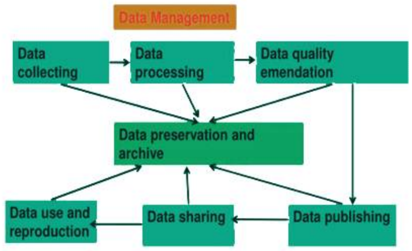

Figure 1: Data Management

II. PRIVACY-PRESERVATION IN DATA PUBLISHING

A typical scenario for data accumulation and publishing is described in Figure 1. In the data amassment phase, the data publisher accumulates data from record owners (e.g., Alice, Bob and peter). In the data publishing phase, the data publisher releases the accumulated data to a data miner or to the public, called the data recipient, who will then conduct data mining on the published data. In this survey, data mining has a broad sense, not obligatorily restricted to pattern mining or model building.For example, a hospital collects data from patients and publishes the patient records to an external medical center. Another example we could site here is financial institutions like banks collect data from customers and publishes the customer record.

We examined the work of (Xiao and Tao, 2006) on the publication of sensitive data using generalization, the most popular anonymizationmethodology in the literature.Abou-el-elaAbdouHussien et.al. (2013).PPDP focuses on techniques for publishing data, not techniques for data mining. This level of involvement is not expected of the data holder in PPDP who usually is not an expert in data mining.

Yan Zhao et,al (2009).At present, privacy preserving in data publishing is at the stage of development.VishwasChavan et.al. (2013).Publishing data should be seen as a necessary step in the publishing process.JiaxingQu et.al (2017). Most of current privacy-preserving schemes which focus on sensitive data sharing issues are dependent on anonymization techniques.ArchanaTomar et.al (2011) In order to solve the privacy preserving problem of association rule in centralized database, before publishing database we should hide the privacy or the sensitive information pattern of the database

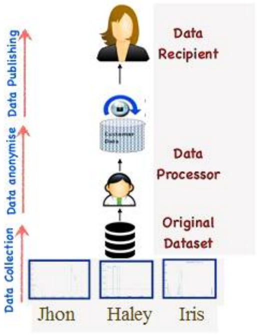

Figure 2: An Overview of Privacy-Preservation in Data Publishing

Bee-Chung Chen et.al(2009) has worked and explained that given a data set, privacy-preservation in data publishing can be intuitively thought of as a game among four parties:

- **Data user** could be member of library, who is interested to use the data or information in library.
- **Adversary**is person who wants to derive private information from the data.

- **Data publisher**, who collects the data and wants to release the data in a way that satisfies the data user's need but also prevents the adversary from obtaining private information about the individuals in the data.
- **Individuals**, who collect the data from the data publisher.

In some cases, the individuals agree with the data publisher's privacy policy, trust the data publisher and give the data publisher all the requested information. In these cases, it is the data publisher's responsibility to ensure privacy preservation. In other cases, the individuals do not trust the data publisher and want to make sure that the data publisher cannot precisely identify their sensitive information. There is a fundamental tradeoff between privacy and utility. At one extreme, the data publisher may release nothing so that privacy is perfectly preserved; however, no one is able to use the data. At the other extreme, the data publisher may release the data set without any modification so that data utility can be maximized; however, no privacy protection is provided. For the data publisher to release useful data in a way that preserves privacy, the following three components need to be defined.

Sanitization Mechanism: - A sanitization mechanism sanitizes the data set by making the data less precise. This mechanism defines the space of possible "snapshots" of the original data set that are considered as candidates for release. We call such a snapshot a release candidate. Generalization is an example sanitization mechanism.

Privacy criterion: - Given a release candidate, the privacy criterion defines whether the release candidate is safe for release or not. K-Anonymity is an example privacy criterion. **Utility metric**: Given a release candidate, the utility metric quantifies the utility of the release candidate (equivalently, the information loss due to the sanitization process).

Given the above three components, one approach to privacy preserving data publishing is to publish the most useful release candidate that satisfies the privacy criterion.

III. KDD PROCESS

Harish Barapatre et.al (2016) narrated that in the present scenario in our day to day life, development in data mining has become very much popular. But, growing popularity and development in data mining technologies brings serious threat to the security of individual's sensitive information. To avoid access to one's sensitive information, Privacy Preserving Data Mining (PPDM) is being suggested. In this technique, data gets modified in order to secure one's sensitive information. PPDM technique is mainly focused on how privacy is maintained at Data Mining. However, Sensitive information can be retrieved at Data Collection, Data Publishing and Information Delivering processes. In this Paper, author has briefly discussed on the Privacy Preservation in Data Mining with respect to user such as Data Provider, Data Collector, Data Miner and Decision Maker. Author has discussed about privacy concerns related to each user with the approach of game theory.

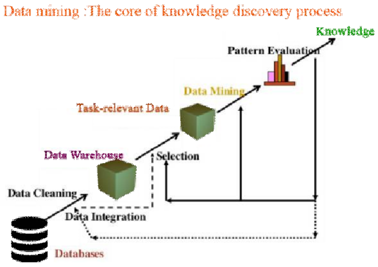

Figure 3: KKD Process

To retrieve useful information from data, following steps are followed:

Step 1: Data Processing. It includes basic operations like Data selection, Data cleaning and Data integration. Data selection process retrieves the data relevant to KDD task from database. Data cleaning process is used to remove noise and inconsistent data and Data Integration process combines data from multiple sources.

Step 2: Data Transformation. Data Transformation performs feature selection and feature transformation.

Step 3: Data Mining. Data mining process extracts the useful data from large amount of data.
Step 4: Pattern evaluation and presentation. It includes basic operations like identifying the truly interesting patterns which represent knowledge and presenting the mined knowledge in an easy-tounderstand fashion.

IV. PPDP VERSES PPDM

❖PPDP focuses on techniques for publishing data, not techniques for data mining. In fact, it is expected that standard data mining techniques are applied on the published data. In contrast, the data holder in PPDM needs to randomize the data in such a way that data mining results can be recovered from the randomized data. To do so, the data holder must understand the data mining tasks and algorithms involved. This level of involvement is not expected of the data holder in PPDP who usually is not an expert in data mining.

❖We develop a personalized filtering scheme with privacy preservation Kuan Zhang(2015).Both randomization and encryption do not preserve the truthfulness of values at the record level; therefore, the released data are basically meaningless to the recipients. In such a case, the data holder in PPDM may consider releasing the data mining results rather than the scrambled data.

❖PPDP primarily “anonymizes” the data by hiding the identity of record owners, whereas PPDM seeks to directly hide the sensitive data. Excellent surveys and books in randomization and cryptographic techniques for PPDM can be found in the existing literature.Abou-el-elaAbdouHussien et.al (2013).

V. DATA CLASSIFICATION

Data classification is broadly defined as the process of organizing data by relevant categories so that it may be used and protected more efficiently. The classification process not only makes data easier to locate and retrieve – data classification is of particular importance when it comes to risk management, compliance, and data security. Data classification involves tagging data, which makes it easily searchable and traceable. It also eliminates multiple duplications of data, which can reduce storage and backup costs, as well as speed up the search process.

Reasons for Data Classification

Data classification is carried out for a variety of purposes, one of the most common being a process that supports data security initiatives. But data may be classified for a number of reasons, including ease of access, to comply with regulatory requirements, and to meet various other business or personal objectives. In some cases, data classification is a regulatory requirement, as data must be searchable and retrievable within specified timeframes. For the purposes of data security, data classification is a useful tactic that facilitates proper security responses based on the type of data being retrieved, transmitted, or copied.

Types of Data Classification

Data classification often involves a multitude of tags and labels, defining the type of data, confidentiality, and its integrity. Availability is also sometimes considered in data classification processes.Data’s level of sensitivity is often classified based on varying levels of importance or confidentiality, which correlates to the security measures in place to protect each classification level. For example, an organization may classify data as:Restricted, Private and Public. In this instance, public data represents the least-sensitive data with the lowest security requirements, while restricted data is in the highest security classification and represents the most sensitive data. This type of data classification is often the starting point for many enterprises, followed by additional identification and tagging procedures that label data based on its relevance to the enterprise, quality, and other classifications.

Restricted data is super sensitive information. Restricted data is linked with "notice-triggering" action. It is mandatory to notify people if there has been unauthorized access or disclosure of this information. Leaks of this type of information can lead to identity theft, news coverage/publicity, and reputational damage and costs to the institution. Examples of this could be Social Security Number (SSN), driver's license/state ID numbers, financial account numbers, credit card numbers, personal medical and medical insurance information, and passwords. Restricted data requires highest level of security, often driven by legal and regulatory requirements and penalties. Individual education is initiation about storing secured data. It is advised not to store on mobile devices, posting on line, in cloud services , sharing to unauthorized people etc.

Private data is moderately sensitive information. It is classified as non-notice-triggering category. This information needs to be protected from unauthorized access. Some examples could be home address and phone, birth date, gender, religious or sexual orientation, and other non-RD personal information; student records, grades, evaluations, letters of recommendation; sensitive research. Proceeding with little caution is acceptable for this type of data.

Public is completely non-sensitive information. Examples could be directory information (name, campus email address, department, etc.) course catalog info, public web pages.It's acceptable to share non-confidential information with others and it can be posted online.

Data Classification Process

The process of data classification can be a complex and cumbersome process, unless automated systems are used to streamline the process. Still, an enterprise must determine the categories and criteria that will be used to classify data, understand and define its objectives, outline the roles and responsibilities of employees in maintaining proper data classification protocols, and implement security standards that correspond with data categories and tags. When done correctly, this process will provide employees and third parties involved in the storage, transmission, or retrieval of data with a framework within which to operate.

Policies and procedures should be well-defined, considerate of the security requirements (or confidentiality) of data types, and straightforward enough that policies are easily interpreted by employees to promote compliance. For instance, each category should include information about the types of data classified as such, security considerations with rules for retrieving, transmitting, and storing data, clear examples, and potential risks associated with a breach of security policies.

The data classification process goes far beyond making information easy to find. Data classification is necessary to enable modern enterprises to make sense of the vast amounts of data available at any given moment. Data classification provides a clear picture of the data within the organization's control and an understanding of where data is stored, how it's most easily accessed, and how data is best protected from potential security risks. Data classification, once implemented, provides an organized information framework that facilitates more adequate data protection measures and promotes employee compliance with security policies.

VI. CLASSIFICATION OF SENSITIVE DATA

Classification of data is done to find the accuracy, sensitivity and specificity percentage of the individual. In this research classification process is used to collect the sensitive and non-sensitive data separately where sensitive data should be highly protected. Without the knowledge of user data can't be published and this is to protect the sensitive data from unknown person. For privacy protecting of data four levels of classifications is planned. They are namely **Restricted Sensitivity, High Sensitivity, Moderate Sensitivity and Low Sensitivity**. This sensitivity data classification is taking into account socio- economic and psychological factors and classifying individual's privacy preference.

Restricted Sensitivity:-Disclosure could cause severe harm to individuals, including exposure to criminal and civil liability. Has the most stringent legal or regulatory requirements and requires the most prescriptive security controls.Legal and/or compliance regime may require assessment or certification by an external, third party. Examples are credit card numbers (PCI) and FISMA

High Sensitivity:- Disclosure could cause significant harm to individuals. Examples could be IT security information, Social Security number, Bank Account numbers, medical information etc.

Moderate Sensitivity:-Disclosure could cause limited harm to individuals. Examples could be employee record, building plans and associated information, immigration documents, intellectual or other proprietary property etc.

Low Sensitivity:-Encompasses public information and data for which disclosure poses little to no risk to individuals. Examples could be public websites, information of public domain.

Examples:

PsychologyFactors

Name	Motivation	Perception	Learning	Attitude and beliefs
Ram	**4**	**1**	**1**	**1**
Nisha	**2**	**2**	**2**	**4**
Ravi	**3**	**4**	**3**	**3**
Priya	**4**	**3**	**4**	**3**

Personal Factors

Name	Age	Caste	Religion	Income	Occupation	Marital Status	Parents Details	Parents Occupation	Contact Number
Meena	4	3	4	3	2	4	2	2	1
vicky	3	4	4	2	2	3	1	1	1
ramya	4	4	4	1	3	3	3	2	1
ramesh	4	4	4	3	2	4	1	1	1

Economic Factors

Name	Salary	Status	Savings	Family Income	Liquid Assets	Income Expect ions	Consumer Credit
Anu	**1**	**1**	**1**	**3**	**1**	**4**	**4**
Divya	**2**	**2**	**2**	**2**	**2**	**2**	**2**
Syed	**4**	**3**	**1**	**4**	**3**	**1**	**1**
Siva	**4**	**1**	**3**	**1**	**1**	**3**	**4**

Social Factors

Name	Name	Age	Gender	Location	Place	Address
Restricted	**1**	**1**	**3**	**1**	**1**	**1**
High	**1**	**2**	**2**	**1**	**3**	**1**
Moderate	**1**	**3**	**3**	**2**	**3**	**1**
Low	**2**	**2**	**3**	**3**	**3**	**1**

The above data classification forms a foundation and the basis for deriving personalised privacy preservation in data publishing.

K – Anonymity

LuoYongchenget.el (2009)K-anonymization techniques have been the focus of intense research in the last few years.Inorder to ensure anonymization of data while at the same time minimizing the information loss resulting from data modifications.K-anonymity model can obviate publishing data from disclosing privacy efficaciously and efficiently. Due to the uneven distribution of the sensitive data, mundane k-anonymization method cannot guarantee each tuple slaking the personalized privacy requisite of its data owner albeit the publishing table has been slaked k-anonymity constraint. JinlingSong (2015) analyzed the reason which k-anonymity table failed to slake personalized privacy requisite and then correlated to the degree of Sensitive Values, Leakage Amassment, privacy disclosure metric and data quality metrics.A. K. Ilavarasi et.al. (2013) Anonymization reduces the risk of identity disclosure whereas the data remains still realistic.Arik Friedman et.al (2007) extended the definitions of k-anonymity and used them to prove that a given data mining model does not violate the k-anonymity of the individuals represented in the learning examples. His extension provided a tool that measures the amount of anonymity retained during data mining. He showed that his model can be applied to various data mining problems, such as classification, association rule mining and clustering. The author has described two data mining algorithms which exploit the extension to guarantee that they will generate only k-anonymous output, and provide experimental results for one of them. Finally, it was shown that their method contributes new and efficient ways to anonymize data and to preserve patterns during anonymization.David J. Martin et.al (2007)k-anonymity does not adequately protect the privacy of an individual.

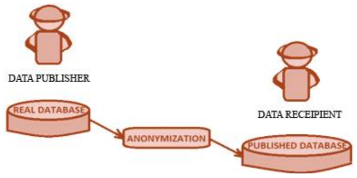

Figure 4: ANONYMIZED DATA TABLE FOR PUBLISHED DB

We consider tabular data where each row is an individual, and the columns (attributes) are labeled “sensitive” or “insensitive”. We want to protect the sensitive attributes.

Table I presents a raw data about matrimony data published, where, every row belongs to the registered individual. After applying, generalization, anonymized data is published in Table II.

Record ID	Bride/Groom	Age	Application ID No	Religion
1	Female	23	56789	Hindu
2	Female	30	56798	Christen
3	Male	28	57689	Muslim
4	Female	21	57698	Hindu

TABLE I: RAW DATA ABOUT MATRIMONY DATA

2−anonymity =⇒

Record ID	Bride/Groom	Age	Application ID No	Religion
1	Female	2*	5678*	Hindu
2	Female	3*	5679*	Christen
3	*	2*	576**	Muslim
4	*	2*	576**	Hindu

TABLE II: ANONYMIZED DATA TABLE FOR PUBLICATION

Game Theory

The study of situation involving computing interest, modeled in terms of the strategies, probabilities, actions, gains, and losses of opposing players in a game is game theory.In other words, the study of games to determine the probability of winning, given various strategies.The analysis of competitive situation (or situations of conflict) using mathematical models. Essential terminology in game theory a) The way a game is played depends on strategy-a plan of action before the game begins. b) A solution is the adoption of a strategy that yields a particular outcome. C) Compare “solving “environmental problems with solving an equation. Ariel Dinar et.al (2015).GT has become one of the basic analytical tools for addressing strategic issues in many fields.

Example: Six people go to a restaurant.Each person pays for their own meal-a simple decision problem.Before the meal, every person agrees to split the bill evenly among them.

Figure 5: Game Theory

Game theory provides mathematical tools for the analysis of strategic interactions, with applications to many fields.

Cooperative outcome	**Equilibrium in a game where the players agree to cooperate**
Dominant Strategy	A dominant strategy is one where a single strategy is best for a player regardless of what strategy other players in the game decide to use.
Nash Equilibrium	Any situation where all participants in a game are pursuing their best possible strategy given the strategies of all of the other participants
Tacit collusion	Where firms undertake actions that are likely to minimize a competitive responsive e.g. avoiding price cutting or not attacking each other’s market.
Whistle Blowing	When one or more agents in a collusive agreement report it to the authorities.
Zero sum game	An economic transaction in which whatever is gained by one party must be lost by the other.

Table III: Key Concepts – Game Theory

Nash Equilibrium

Nash equilibrium is an important idea in game theory.It describes any situation where all of the participants in a game are pursuing their best possible strategy given the strategies of all of the other participants.In a Nash equilibrium, the outcome of a game the occurs is when player the A takes the best possible action given the action of player B, and player B takes the best possible action given the action of player A.

Column

Row

	a	**b**
a	**1,2**	**0,1**
b	**2,1**	**1,0**

Table IV: Example of Nash Equilibrium

(b,a) is a Nash equilibrium

To Prove this:

- Given that column is playing a, row’s best response is b.
- Given that row is playing a, column’s best response is a.

VII. CONCLUSION

In this paper, classification of data based on sensitivity is done. This forms the basis for a theoretical approach of personalized privacy preservation in data publishing. The goal of this work is to implement a practicalpersonalized privacy preserving framework to keep privacy of an individual as per the personalized preference of the individual. The core benefit of this work is to promote data sharing with personalized privacy protection.Based on the classification of sensitivity, a practical implementation is plannedusing Game Theory with concept of Nash Equilibrium for achieving personalized privacy preservation in data publishing. This is planned to be implemented on specific area like financial sector where personalized privacy preservations is considered to be important and is under the classification of restricted sensitivity.

REFERENCES

1. Arik Friedmanetr.al.(2006)k-Anonymous Decision Tree Induction, Appeared in The 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), Berlin, September 2006.
2. Abou-el-elaAbdouHussien et al.(2013)Based Privacy-Preserving: A Survey for Data Mining and DataJournal of Information Security, 2013, 4, 101-112 Published Online April 2013.
3. Bee-Chung Chen et.al.(2009)Privacy-Preserving Data Publishing Foundations and Trends in Databases Vol. 2, Nos. 1–2 (2009) 1–167 c 2009 B.-C. Chen, D. Kifer, K. Lefebvre and A. Machanavajjhala DOI: 10.1561/1900000008.
4. BENJAMIN C. M. FUNG et.al. (2010)Privacy-Preserving Data Publishing: A Survey of Recent Developments. DOI = 10.1145/1749603.1749605.
5. GeorgiosChalkiadakis et.al.(2012)Computational Aspects of Cooperative Game www.morganclaypool.comISBN: 9781608456529 paperback ISBN: 9781608456536 eBook DOI 10.2200/S00355ED1V01Y201107AIM016.
6. Harish Barapatre et.al.(2016)Privacy Preserving using data Anonymization technique on Hadoop Volume 5, Issue 4, April 2016.
7. J.R.Jayapriya et.al.(2013)Game Theory Approach for Identity Crime (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 12, Issue 3 (Jul. - Aug. 2013).
8. Jinling Song et.al. (2015)The K-Anonymization Method Satisfying Personalized Privacy Preservation VOL. 46, 2015 AIDIC ServiziS.r.l. ISBN 978-88-95608-37-2; ISSN 2283-9216.
9. V.S. Lakshmanan et.al. (2000) On DominationGame Analysis for Microeconomic Data Mining. °c 20YY ACM 0000-0000/20YY/0000-0001 $5.00 ACM Journal Name, Vol. V, No. N, Month 2000.
10. LEI XU et.al.(2014)Information Security in Big Data: Privacy and Data Mining Digital Object Identifier 10.1109/ACCESS.2014.2362522.
11. Raymond Chi-Wing Wong et.al.(2009) Anonymization with Worst-Case Distribution-Based Background.
12. Xiaokui Xiao Yufei Tao et.al. (2006)Personalized Privacy Preservation, SIGMOD 2006, June 27–29, 2006, Chicago, Illinois, USA.
13. Xiaohui Liang et.al. (2015)A Personalized Fine-Grained Spam Filtering Scheme With Privacy Preservation in Mobile Social Networks. IEEE transactions on computational social systems Vol 2, No. 2 September 2015.
14. Xiao tong Wu et.al. (2016)Game Theory Based Correlated Privacy Preserving Analysis in BigDataDOI10.1109/TBDATA.2017.2701817,http://www.ieee.org/publications_standards/publicatiions
15. Abou-el-elaAbdouHussien , NerminHamza, Hesham A. Hefny. () “Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing”. Journal of Information Security, 2013, 4, 101-112 http://dx.doi.org/10.4236/jis.2013.42012 Published Online April 2013 (http://www.scirp.org/journal/jis)
16. A N K Zaman () “Privacy Preserving Data Publishing: A Classification Perspective”. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No.9, 2014.
17. AshwinMachanavajjhala (2009) “Data Publishing against Realistic Adversaries”. VLDB ‘09, August 24-28, 2009, Lyon, France Copyright 2009 VLDB Endowment, ACM 000-0-00000-000-0/00/00.
18. VishwasChavan, LyubomirPenev, and Donald Hobern.(2013) “Cultural Change in Data Publishing Is Essential”.June 2013 / Vol. 63 No. 6 .doi:10.1525/bio.2013.63.6.3
19. A.Dinar and M. Hogarth. “Game Theory and Water Resources: Critical Review of its Contributions, Progress and Remaining Challenges. Foundations and Trends R in Microeconomics”, vol. 11, nos. 1–2, pp. 1–139, 2015.
20. David J. Martin Daniel KiferAshwinMachanavajjhala Johannes Gehrke (2007). “Worst-Case Background Knowledge for Privacy-Preserving Data Publishing”.
21. 7 AK. Ilavarasi “A Survey on Privacy Preserving Data Mining Techniques”. ISSN: 1694- 2108 | Vol. 7, No. 1. NOVEMBER 2013.
22. AsmaaHatem Rashid (2015). “Sharing healthcare information based on privacy preservation”http://www.academicjournals.org/SRE.
23. Yan Zhao(2009). A Survey on Privacy Preserving Approaches in Data Publishing. 978-0-7695-3604-0/09 © 2009 IEEE DOI 10.1109/DBTA.2009.149.

24. LuoYongcheng, Le Jiajin and Wang Jian (2009) "Survey of Anonymity Techniques for Privacy Preserving. International Symposium on Computing, Communication, and Control" (ISCCC 2009) Proc .of CSIT vol.1 (2011) © (2011) IACSIT Press, Singapore.
25. JiaxingQu,Guoyin Zhang, and Zhou Fang (2017). "A Context-Aware Location Privacy-Preserving Scheme in Location Sharing Service". Discrete Dynamics in Nature and Society Volume 2017, Article ID 6814832, 11 pages https://doi.org/10.1155/2017/6814832
26. Amardeep Singh (2014) "Privacy Preserving Techniques in Social Networks Data Publishing"-A Review.. International Journal of Computer Applications (0975 – 8887) Volume 87 – No.15, February 2014.
27. Amardeep Singh (2014) "A Comprehensive Survey of Privacy Preserving Algorithm of Association Rule Mining in Centralized Database". International Journal of Computer Applications (0975 – 8887) Volume 16– No.5, February 2011.
28. "A Personalized Fine-Grained Spam Filtering Scheme With Privacy Preservation in Mobile Social Networks"(2015). IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, VOL. 2, NO. 3, SEPTEMBER 2015.
29. SlawomirGoryczka. (2011). "m-Privacy for Collaborative Data Publishing".
30. Xiao X, Tao Y (2006). "Personalized privacy preservation".