Identification of K- Tuples using K-Anonymity Algorithm to the Watermarking of Social Network Database

Rajneeshkaur Bedi
Associate Professor
MIT COE,
Pune, India

Anjali Mahajan, PhD
Professor and Head,
Government Polytechnic Institute
Nagpur, India

ABSTRACT

Nowadays digital data is increasing and securing these assets is vital. Generating and distribution of digital content via Internet has raised many opportunity for good and bad reasons. Many cases on piracy and privacy are registered for falsification, tampering or forgery on digital data like database, multimedia, document etc. And most important is to protect personal privacy. In view of, these two problems of piracy and privacy our research focused in identification of k-tuples using k-anonyminity algorithm and watermarking them. Here, privacy preserved K-anonymity algorithm is modified and developed to identify the sensitive attributes to transfer to other parties after solving the privacy issues. Then, the sensitive attribute is modified to preserve the privacy in a way that to handle the piracy of database. This provides additional advantage of copyright protection along with privacy preservation. After exact identification the sensitive tuples by the k-anonymity algorithm, the tuples will be watermarked based on our embedding procedure.

General Terms

Social Networking, watermarking, anonymization

Keywords

Privacy, piracy, k- anonymity algorithm, watermarking

1. INTRODUCTION

Social network collects lots of data from the user and over the period huge data is gathered in various forms. This data is viewed as a collection of entities and links between them [6]. The entities are nothing but an individual who are connected or linked with others through some relationship, comments or likes. Analytical studies on social network leads to uncover various patterns like finding the popularity, influence of comments, detect fraud and their linkage. Technological advances researchers applied social network analysis to study disease transmission, functioning of computer networks behaviour of physical and biological system. However, such data is facing two undesirable outcomes [8]. Lots of vulnerabilities are present in the released dataset which is exploited by the intruders, so the biggest challenge for the research community is to address the protection of social network data [7]. Social network data analysis is very fruitful for further studies only concern is to protect the privacy of an individual's sensitive and confidential data. The privacy rules and regulation for sharing and accessing personal data exist but organization need to comply the privacy preference of the data owner [10][15]. Researchers need person specific data from various organizations like medical institution, financial, public health etc for further analysis. The challenge is to hide the identity of an individual while sharing the data. The first suggested method to achieve this is to adhere K-anonymity [11]. In this approach individual's sensitive attribute value is modified with generalization and suppression technique to protect privacy. This method ensures at least k-1 record in released dataset is indistinguishable. The k factor is the anonymization degree and also shows desired level of privacy. Still it's difficult to implement due to nature of released data and its access by the traditional approach of k-anonymity [4][13].

Another issue which researchers are facing is to handle copyright of electronic data from the attackers. The most acceptable way to protect copyright of such electronic data is watermarking. Watermarking protects the ownership of an individual's data with its unique mark. To prove individuals ownership is difficult without watermarking. A watermark consists of a unique mark which is encoded in original data. This mark is used to resolve or prove the ownership of data. Nowadays watermarking become a hot topic for relational database, text, document, multimedia like audio, video and images [2][12][14].

This paper gives the identification of K- tuples using K-anonymity algorithm to the watermarking of social network database. The contributions of the proposed approach are,

- A detailed Study of the different K-anonymity Algorithm schemes
- Designed and developed an efficient Watermarking technique for structured database without affecting the privacy.
- An Analysis on performance of proposed algorithm in identification of K- tuples and the strength of watermarking of social network database is conducted

Rest of the paper is organized as, the 2nd section plots a short review of the recent research and the 3rd section plots the detailed description about the proposed approach and information loss. The 4th section includes the performance analysis and comparative analysis of the proposed approach. With the 5th section conclude the paper.

2. REVIEW OF LITERATURE

This section, list of research plotted some regarding the k-anonymity and information loss. The most of the researches are trending towards optimizing k-values. Most of the results are concentrating on improving the efficiency of the privacy preservation.

Latanya Sweeney [11] has included a k- anonymity algorithm for privacy protection with the use of generalization and suppression. The k-anonymity method results in such a way that each individual in the release data is distinct than other by at least k-1. The k-anonymity method is also appreciated for re-identification attack after studied properly when all the policies are followed properly. To deal with real world system Datafly and μ-Argus the k-anonymity protection model is used along with them. The study shows that Datafly and µ-Argus, always not satisfied k-anonymity, but generalizations in this way also sometimes result in k-minimal distortions.

Specialized applications like healthcare are privacy sensitive which need special efforts for

privacy protection. Generally such type of applications data is anonymized, but the issue comes when new user joins. To address these problems either re-anonymization for the entire dataset is done or incremental method of anonymization is deployed. The re-anonymization method is time consuming; while incremental method suffers from scalability and efficiency issue for huge dataset. Zhang et al. [] proposed indexing method on quasi-identifier to handle privacy protection efficiently. Systematically quasi-identifier is indexed in distributed way using locality sensitive hash function to improve efficiency. This increases number of data nodes that a QI-group link across dataset which was reduced considerably with high probability.

K-anonymity concept inspired to K. P. Kaliyamurthie et al [5] to define it for designing a security framework in wireless sensor networks (WSN). This framework consists of two levels of privacy using semi trusted sink and deeper level privacy. In this approach encryption and generalization is used. In order to save energy, generalization is used to reduce the size of data transmitted by sacrificing information loss. To overcome or reduce information loss due to generalization a method of bottom up clustering is used. This selects some data portion for encryption which is actually causing more information loss. This achieve good amount of energy saving keeping information loss in control.

A survey paper by Rashmi A. Zilpelwar et al [1] highlighted the importance of rising demands of Social networking sites and its application with the associated risk. As, most of the users are not aware of the threats involved in using SNS or their application landed up in mess of personal privacy. They examined the security and privacy provided with existing SNS and brought out risks and challenges which user may face regarding personal sensitive data to avoid potential loss of private and personal information. Automatically lot of data is collected without the user knowledge in Social networking sites which become a weapon for the attacker. They listed out all the potential attack and surprisingly all the social networking site are venerable to these attacks.

Watermarking the relational database for data authentication and integrity is an important aspect which is highlighted by Anita Thengade et al [3] for text data, and to achieve it they proposed a novel approach of using mathematical concept of Eigen values. The original database along with watermark key can be used for verifying the integrity of database. Their method draw a secret key by using vowels, consonants and special character and evaluated as Eigen matrix. The matrix is constructed using tuple – relation matrix for each record independently. The secret key is used to insert watermark in low impact non-numeric attribute of a tuple. Authentication is proved after successful watermark detection. The experimental result showed that the new approach is effective and robust against different forms of malicious attacks as well as benign updates to the data.

Another approach for preserving privacy is proposed by V.Thavavel a nd S.Sivakumar [9] in unstructured data environment. This framework is for meta-data using hiding technique with a distributed mechanism. They worked on text data type. Text processing on document using words occurrence and similarity test is used. A novel approach of semantic and syntactic way to watermark is proposed by Purva Gujarathi et al [2]. The algorithm was based on the concept of predefined signals for ASCII characters and abbreviation of words. A secret key has been generated by using predefined signals. To embed a watermark they used two options one is the concept of abbreviations for words as syntactic approach, and another is by using ASCII value of some digits of secret key if former technique was not possible. Detection of watermark leads to the authentication and integrity to data. Experimental result shown that, their approach was robust and secure against the various malicious attacks.

3. IDENTIFICATION OF K- TUPLES USING K-ANONYMITY ALGORITHM TO THE WATERMARKING OF SOCIAL NETWORK DATABASE

3.1 Privacy preservation through k-anonymity

Data privacy is currently deals with the implementation of K-anonymity approach, which ensure that each record in table is identical to atleast (k-1) other records with respect to sensitive attributes called quasi identifier. Consider table 1 a sample social network data and table 2 after applying K-anonymity with generalization and suppression. The suppression type anonymity is mostly like a deletion of information, while generalizations simply generalizes the data and convey a simple part of information. But the point is to conserve most of the information by preserving the anonymity. Our approach deals in decision making on k value so as to reduce the information loss.

Table.1. sample social network data

Sr. No	Name	Age	Place	Sex	Ph. No	Occupation
1	Deeepak	31	Calicut	M	4312	Accountant
2	Nile	34	Amravati	M	4315	Assistant
3	Smita	29	Delhi	F	4311	Assistant
4	Ketaki	32	Banglore	F	4318	Manager

Table.2. 3- anonymity with generalization and suppression

Sr. No	Name	Age	Place	Sex	Phn. No	Occupation
1	Deeepak	25-35	Calicut	M	43**	XX
2	Nile	25-35	Amravati	M	43**	XX
3	Smita	25-35	Delhi	F	43**	XX
4	Ketaki	25-35	Banglore	F	43**	XX

So, it incorporate the data mining technique, namely data clustering. The clustering give a partial solution to information loss, that needs to specifically design the clustering algorithm to fetch the k value for k anonymity. The information loss over a data mainly associated with the cluster size and attributes regarding a cluster.

3.2 Information loss

Information is lost in anonymization via suppression and generalization. But the question is how much is lost. Various methods are proposed by researcher, this paper follows the measurement given by Byun et.al [3] for an efficient systematic clustering method. Referring the formula given by Byun et.al [3] for the amount of information loss due to generalizing x which is denoted by IL(x) is given as:

$$IL(x) = |x| \left(\sum_{i=1}^{r} \frac{N_{i\max} - N_{i\min}}{\eta N_{i\max} - \eta N_{i\min}} + \sum_{j=1}^{s} \frac{H(\Lambda(\cup_{C_j}))}{H(\tau_{C_j})} \right)$$

and the total information loss as:

$$IL(total) = \sum IL(x_i)$$

where, quasi identifier in numeric identifier N1 to Nr, categorical identifier C1 to Cs, τCi be the taxonomy tree defined for the domain of Ci, | x| is the number of records in x, τ(∪Cj) is the sub-tree rooted at the lowest common ancestor of every value in ∪Cj and H(τ) is the height of taxonomy tree τ.

The main objective of the proposed approach is to minimize the information loss regarding a cluster. The proposed approach adopted a different way to incorporate relevant data into the cluster to minimize the information loss. The formation of cluster is the key to determine the k values in relevant get efficient anonymity to the data provided. So, proposed method has adopted a modified systematic clustering algorithm to provide k-anonymity with controlled information loss.

3.3 Modified systematic clustering for k-anonymity

The clustering process is a useful medium to determine the k-value for k-anonymity process. So in our proposed approach, is associating a modified systematic clustering approach to determine the k- value. A systematic clustering basically means a procedure to cluster the data with different k-values and finally, k value will be selected from the cluster that possesses less information loss. Start to form cluster with initial k value and see to minimize intra cluster and maximize inter cluster distance. In systematic clustering method concentrate to minimize intra cluster distance and maximize the value of anonymity. The inter cluster distance is defined in the time of cluster formation as the elements in the cluster should be close in order to minimize the inter cluster distance. So, a define radius for each element in the data and the elements which come under the radius of an element will form as a cluster.

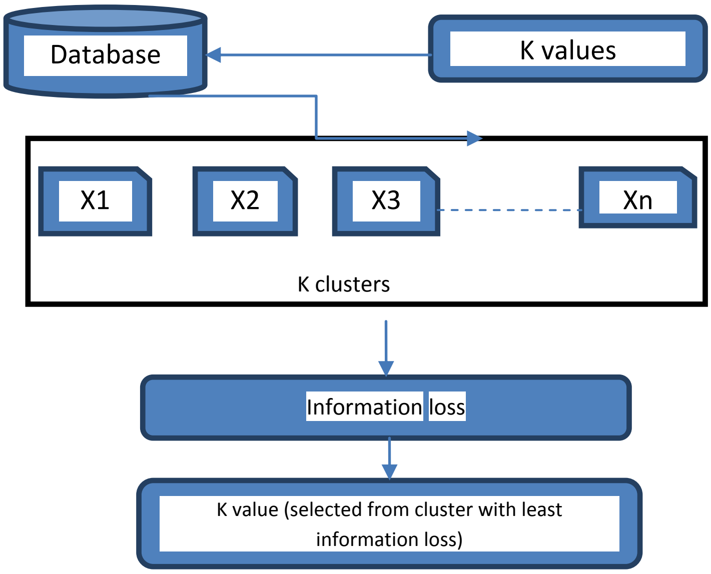

Fig.2. systematic clustering

Now, move on to the clustering process.

```
Algorithm systematic clustering

Input: records from social network
Output: k value, clusters

Step1. Select dataset D,
Step2. Define cluster size Z, radius e
Step3. Initiate k value
Step4. Form cluster for ri records
Step4. If $z = k$
Stop adding records
Else
Add records with e
Step5. Consider the formed clusters xi
Step6. Calculate information loss, IL (x)
Step7.simiarly consider all the clusters and ILs
Step8. Select cluster xi, where IL (xi) = low
Step9. If any records are need to process add them into xi

Step10. Select k value form xi

Step11. End
```

3.4 Piracy preservation through embedding secret message

The social network data may contain lot of vital information about the particular user, so ensuring privacy is not alone give the user anonymity. Thus, ended up with a solution of encapsulating secret data into user's name. This will prevent users to copy or extract other user information with authorization. In the proposed approach, use of text based cryptography to protect the vital information. The text based cryptography is used because; use of the field "name" from the records of user. Embedding particular secret message to the users name to protect the record from unauthorized access.

3.4.1 Encryption

The encryption process is based on basic text steganography and use an incremental or decrement scheme. Consider the a tuple t = [a1, a2,…,an] here, t is the tuple and "a" represents the attributes on the tuple. Among the n attributes (n-k) attributes are not gone through any anonymity process and the one particular attribute which can't be used anonymized is "name" attribute. So embed the secret message to the field name. The encryption scheme on field name is explained below with sample,

101	Rex	25-35	California	M	43**	XX

Here, name = Rex

So now apply the encryption letter by letter on the name Rex. The encryption process will generate a key after embedding the message. So let us consider the message "TEA" and embed this message into it by considering the English alphabet. Now consider each letter is assigned a numeric value based on their position, consider the following table.

Table.1. numeric value assigned for each letter

*	1-a	2-b	3-c
4-D	5-e	6-f	7-g
8-H	9-I	10-j	11-k
12-L	13-m	14-n	15-o
16-P	17-q	18-r	19-s
20-T	21-u	22-v	23-w
24-X	25-y	26-z	#

Now, to embed "TEA" on "REX", so initially select position of R and then position of T with respect to position of R. Similarly each letter is processed with respect to the corresponding letters. Position of R= 18 , so initial key value for embedding T will be 2, which is obtained by, position (R)-position (T)= 2. In order to identify the letter sequence, add a prefix in front of the key of each letter. Thus the key value generated for encrypting message TEA on name REX will be, T on R give 2, E on E give 0,

A on X give 3, this is because if the difference is more than 13, just consider forward increment. I.e. X is 3 position away from A based on the table 1. In order to avoid easy decryption mark a bar(-) over the forward incremented values.

So, TEA on REX gives, 203¯, by adding the sequential prefix get the key as ‚T12T20T33¯. The decryption of this key will be easy and so do a single value decrement on each letter's key. Finally, the decrypting key for REX will be,

TEA on REX ➔$T11T226T3\bar{2}$

In the above sample, the total number of letters embedding message and name is same, but in the case of difference in total number, divide the total number of embedding message with total number of letters in the name. Then each letter in the name is given equal number of letters in the embedding message. For example, if it need to embed VENICE in REX, the embedding will take place as follows,

Table.2. Dividing letters to embed

R		E		X	
V	E	N	I	C	E

In this way encrypt secret message to each name in the social network record and the receiver, who has the authorisation can access the information by providing the key.

4. EXPERIMENTAL ANALYSIS

4.1 Experimental setup and dataset description

The main idea of this experiment is to check the recital of our approach in terms of data quality and computational efficiency. The experimentation of the proposed approach is carried out on a system running with i5 processor, 4 GB RAM and 500HDD. The programs are implemented using the java programming language running with JDK 1.7.0. The experimentations are conducted on dataset developed manually based on the information from social working site called LinkedIn. The collected data contains nine different attributes and 1000 users profile are selected for the experimental purpose. The each row is considered as tuple and attribute values are listed in columns.

4.2 Performance Evaluation

The performance of the proposed approach is evaluated based on dataset collected. The evaluation parameters users for the performance analysis are execution time and total information loss. The information loss is calculated based on the equation plotted in the section 3 and time for execution is calculated based on the time in which an operation on the dataset completes. The performance evaluation is carried out for different k values. The k values range from 2 to 5 as it has limited number of accessible attributes in the dataset. The responses of the dataset for different k values are plotted on the below graph,

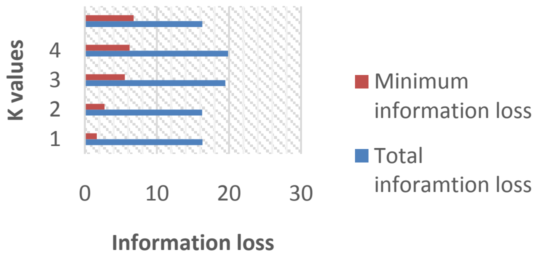

Fig.3. information loss over different k values

The figure 3 represents the information loss over different k-anonymity values. The graphs are plotted based on the total information loss per cluster and minimum information on the clusters.The analysis from the graph showed that, as the number of k-value increases, the minimum value of information loss increases accordingly. The minimum information loss obtained for different k-values increases proportionally. Then select a k-value from a cluster, which maximize the k value and minimize the information loss. Thus, considering the average information loss values, one can easily select the preferred cluster and hence the k- value.

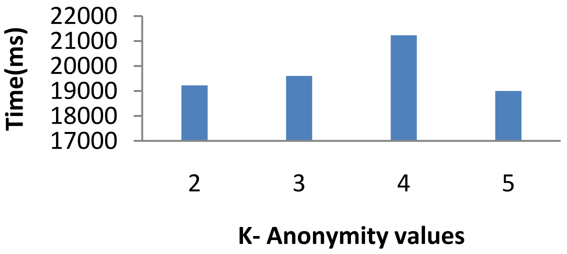

Fig.4. Execution time

The figure 4 shows the time taken for the dataset to execute for different credentials of k- values. The figure shows that, the execution time is not in a regular manner as one can see for k value 2, the execution time near to 19000 ms while it is close to 22000ms for k value 4. It comes down to 19000 again for k value 5. The difference in time occurs because of the processing of categorical attributes in clusters. Different clusters posses' different length of categorical tree, which in turn affect the total execution time.

Table.3. embedding secret message

Text	Encryption Message	Key
HR at Towers Watson	India	1@E8@E16@5@ZE6@E12@E4@E-2@ZE19@E0@E12@E12@ZE18@E13@6@E8@ZE18@E13@E11@Z
HR Ravi	India	E5@Z5@ZE8@ZZZ
Head - HR : TL&amp; D&amp; Corporate Affairs at Barclays Shared Services	India	1@3@6@2@E14@E15@E12@-6@E1@E9@E8@E8@E6@E5@E-3@Z2@E15@11@-2@E-2@E12@4@E6@E8@4@-9@E-9@E5@E0@-3@ZE11@E13@E10@E8@E10@-2@E10@E-6@E15@-6@E-8@E-9@-9@E-8@E0@ZE10@E17@6@E8@E14@2@2@E2@-2@E-6@-2@E5@E-2@E5@E-4@ZE7@-1@E15@E1@E-1@E21@E12@E-3@E9@E12@E-2@E-9@E-8@E5@Z

The table 3 represent the embedding process of the proposed approach. Here select three text data from data field "profession" from the dataset created. The encrypted word "India" in each of the texts and corresponding key is shown in the subsequent column. The encryption procedure is incorporated to the data in order to protect the data from piracy of copying.

4.3 Comparative Analysis

In this section a graph plot gives the comparative analysis of the proposed approach with an existing approach. The comparative analysis is plotted based on the parameters information loss and execution time. The recent work by Md. Enamul Kabir et al [1] has selected for the comparative analysis. The early created dataset is given to both the existing approach and the proposed approach. The responses of both approaches are mapped for detailed analysis.

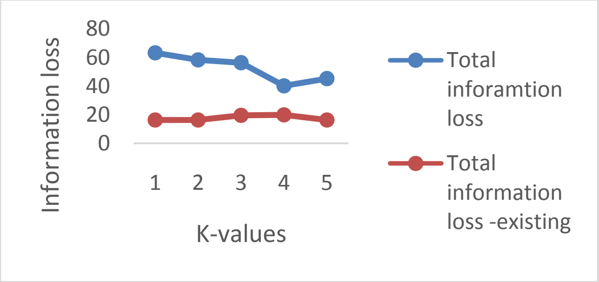

Fig.5. Analysis on information loss

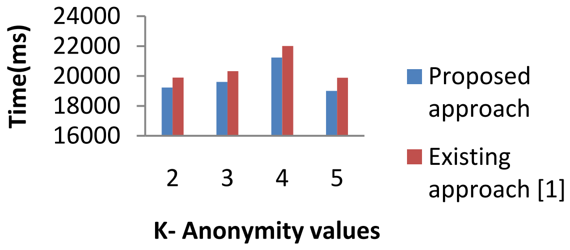

Fig.6. analysis on execution time

The figure 5 and 6 shows the analysis over information loss and execution time by comparing existing approach [1] and proposed approach. The analysis conducted for K-anonymity value ranging from 1 to 5. The graphs clearly states that the proposed approach is up head to the existing approach with less information loss and time for execution. Thus it can state the proposed approach is efficient in handling information loss.

4.4 Cluster-wise Comparative Analysis

The proposed method is an algorithm defined to minimize the information loss and maximizing the privacy of user in the network. One of the prominent method used in the proposed approach is the data clustering method that ensures the information loss and privacy. The proposed method considers two parameters from the clustering method, the inter cluster distance and intra-cluster distance. The objective of the proposed method is to minimize the inter cluster distance and intra cluster distance between the distance. Let us consider the details of inter-cluster distance and intra-cluster distance obtained for the proposed approach by subjecting an input data, which is extracted from the social network.

Table.4. inter cluster distance

Sr. No	Cluster to Cluster	Centroid	Distance between Centroid
1	C1 to C2	1.2 , 2	0.8
2	C2 to C3	2 , 1.1	0.9
3	C3 to C4	1.1 , 2.2	1.1
4	C4 to C5	2.2 , 0	2.2
5	C5 to C6	0 , 1	1
6	C6 to C7	1 , 2.1	1.1
7	C7 to C8	2.1 , 0.2	1.9

The table 4 represents the inter cluster distance between selected cluster that are derived from the input data. The assumption of the proposed approach is that, the lesser the inter cluster distance the lesser will be the information loss. So from the table it can be identified the some clusters are with less difference in distance and some with higher difference. The clusters c1,c2, c3, c5 and c6 are considered by the proposed approach as the threshold is set as 1 or less than 1 as the difference in cluster distances. The lesser inter cluster distance means that those cluster possess similar information as compared to another. So the chances of information loss can be controlled by selecting clusters with lesser inter cluster distance. The proposed approach is also concern about how a cluster is formed. The cluster formation is triggered by intra cluster distances, i.e. if the cluster is tightly associated to the information that are similar, then the loss of information can be controlled and privacy also ensured for a particular individual user. The intra cluster distance is calculated as the difference between centroid of the cluster and other data points between the clusters.

Table.5. intra-cluster distance

Sr. No	Cluster	Avg. intra cluster distance	Total number of elements
1	cluster 1	10303.0	153
2	cluster 2	7540.5	103
3	cluster 3	11245.5	93
4	cluster 4	10715.5	78
5	cluster 5	7904.5	58

The table 5 represents the average intra-cluster distances possessed by clusters formed the input data. The analysis from the table shows that the clusters 1, 2, 3, 5 and 6 possess the least intra-cluster distances. In the case of intra-cluster distance also, the lesser value will do the favour. So it can select the above mentioned clusters for further processing.

5. CONCLUSION

In this paper an approach on identification of k- tuples using k-anonymity algorithm is proposed. The proposed approach develops modified clustering approach to handle the information loss on anonynizind a data. The aim of the proposed approach is to minimize the information loss and maximize anonymity. Through this can ensure more protection on private data of users in a social network. The proposed approach also provides an encryption algorithm for copyright protection of sensitive data in social network. A text steganography based method is used for the above process. Experimental analyses are conducted to analyse the performance of the proposed approach. The proposed approach has provided a minimum of information loss of 1.6667. The comparative analysis stated that, the proposed approach is efficient in identify k-value for anonymity process in the social network. In future, the work can be extended for multiple text field for cryptography and use of native language phonems can be used to generate screte key in watermarking the data.

6. ACKNOWLEDGEMENT

This is the research work associated to the PhD work going on under Sant Gadge Baba Amravati University (SGBAU) and PG Department of Computer Science and Engineering, Amravati. My sincere thanks to the centre Head, Dr. V. M. Thakre and the faculty for their extensive support.

7. REFERENCES

[1] RashmiA.Zilpelwar, RajneeshkaurK.Bedi and Vijay M.Wadhai, "An Overview of Privacy and Security in SNS", International Journal of P2P Network Trends and Technology, vol. 2, no. 1, p p.23-26, 2012.

[2] Purva Gujarathi, Rajneesh Kaur Bedi, Poonam Gundecha and Ashish Kulkarni, "A Unique Approach for Watermarking Non-numeric Relational Database", International Journal of Computer Applications, vol. 36, no.7, p p.9-14, December 2011.

[3] Anita Thengade, RajneeshkaurBedi, and Vijay M.Wadhai, "A New Watermarking Approach for Non-numeric Relational Database", International Journal of Computer Applications, vol. 13, no.7, p p.37-40, January 2011.

[4] Mohammad Rasool Sarrafi Aghdam and Noboru Sonehara, "On enhancing data utility in k-anonymization for data without hierarchical taxonomies", International Journal of Cyber-Security and Digital Forensics, vol. 2, no. 2, p p. 12-22, 2013.

[5] K. P. Kaliyamurthie, D. Parameswari and R. Udayakumar, "k-anonymity Based Privacy Preserving for Data Collection in Wireless Sensor Networks", Indian Journal of Science and Technology, vol 6, no. 5S, p p. 4604-4613, May 2013.

[6] Bruce Kapron, Gautam Srivastava and S. Venkatesh, "Social Network Anonymization via Edge Addition", Advances in Social Networks Analysis and Mining (ASONAM), vol.1, no.8, p p.155-162, july 2011.

[7] Traian Marius Truta and Alina Campan, "K-Anonymization Incremental Maintenance and Optimization Techniques", ACM symposium on Applied computing, vol. 1, no. 4, p p. 380-387, 2007.

[8] Michael Hay, Gerome Miklau and David Jensen, "Anonymizing Social Networks", Security and Privacy, vol.11, p p.173-198, 2009.

[9] V.Thavavel and S.Sivakumar, "A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment", International Journal of Computer Science Issues, vol. 9, no. 2, p p. 434-441,January 2012.

[10] Mohamed Y. Eltabakh, Jalaja Padma, Yasin N. Silva and Elisa Bertino, "Query Processing with K-Anonymity", International Journal of Data Engineering, vol. 3, no. 2, p p. 48-65, 2012.

[11] Latanya Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol.10, no. 5, p p. 571-588, 2002.

[12] Athanasios Nikolaidis and Ioannis Pitas, "Region-Based Image Watermarking", IEEE transactions on image processing, vol. 10, no. 11, p p.1726-1740, november 2001.

[13] Aaron Beach, Mike Gartrell and Richard Han, "Social-K: Real-Time K-Anonymity Guarantees for Social Network Applications", International Association of Computer Science and Information Technology, vol. 28, p p.101-106, 2011.

[14] Nagarjuna. Settipalli and Manjula, "Securing Watermarked-Relational Data by Using Encryption and Decryption", RPN Journal of Systems and Software, vol.1, no.2, p p.70-74, 2011.

[15] Jian, W., L. Yongcheng, J. Shuo and L. Jiajin," A survey on anonymity-based privacy preserving," International Conference on E-Business and Information System Security, p p. 1-4, 2009.

[16] Xuyun Zhang, Chang Liu, Surya Nepal, Jinjun Chen,"An efficient quasi-identifier index based approach for

privacy preservation over incremental data sets on cloud, "Journal of Computer and System Sciences, 2013.

8. AUTHOR PROFILE

Rajneeshkaur Bedi has received B. E. degree in Computer Engineering from Amravati University, India in 1997 and M.Tech degree in Computer Engineering from Pune University, India in 2005. She is a registered Ph.D student of Amravati University. She is working as Assosiate Professor in Computer Engineering at MITCOE, Pune, India from last 12 years. She has more than 15 years of teaching experience. Her research interest includes Data mining, Data Privacy, Natural language processing, cyber forensic, cryptography and Machine learning. She is a life member of ISTE and CSI.

Dr. Anjali R. Mahajan has completed her Ph.D. in Computer Science and Engineering from SGBAU, Amravati, India. Her research areas include data mining, computational intelligence, machine learning, networking and, image processing. She is currently working as Professor and Head of Department of Computer Engineering and Information Technology, at the Government Polytechnic Institute, Nagpur, India. She has more than 20 papers published in various areas in reputed international and national journal . She is a member Machine Intelligence Research Labs, USA among other professional bodies.