K-ANONYMITY: The History of an IDEA

Swagatika Devi

Abstract*— Publishing data about individuals without revealing sensitive information about them is an important problem. In recent years, a new definition of privacy called k-anonymity has gained popularity. In a k-anonymized dataset, each record is indistinguishable from at least k−1 other records with respect to certain "identifying" attributes. In this paper, we discuss the concept of k-anonymity, from its original proposal illustrating its enforcement via generalization and suppression. We also discuss different ways in which generalization and suppressions can be applied to satisfy k- anonymity. By shifting the concept of k-anonymity from data to patterns, we formally characterize the notion of a threat to anonymity in the context of pattern discovery. We provide an overview of the different techniques and how they relate to one another. The individual topics will be covered in sufficient detail to provide the reader with a good reference point. The idea is to provide an overview of the field for a new reader from the perspective of the data mining community.

Index Terms*— K-Anonymity, Generalization, Suppression, Pattern discovery.

I. INTRODUCTION

Privacy Preserving Data Mining, i.e., the analysis of data mining side-effects on privacy, has recently become a key research issue and is receiving a growing attention from the research community ([7],[13]). However, despite such efforts, a common understanding of what is meant by privacy is still missing. This fact has led to the proliferation of many completely different approaches to privacy preserving data mining, all sharing the same generic goal: producing a valid mining model without disclosing private data. As highlighted in [13], the approaches pursued so far leave a privacy question open: do the data mining results themselves violate privacy? Put in other words, do the disclosures of extracted patterns open up the risk of privacy breaches that may reveal sensitive information? In this paper in particular, we focus on individual privacy, which is concerned with the anonymity of individuals. A prototypical application instance is in the medical domain, where the collected data are typically very sensitive, and the kind of privacy usually required is the anonymity of the patients in a survey. Consider a medical institution where the usual hospital activity is coupled with medical research activity. Since physicians are the data collectors and holders, and they already know everything about their patients, they have unrestricted access to the collected information. Therefore, they can perform real mining on all available information using traditional mining tools not necessarily the privacy preserving ones. This way they maximize the outcome of the knowledge discovery process, without any concern about privacy of the patients which are recorded in the data. But the anonymity of patients becomes a key issue when the physicians want to share their discoveries (e.g., association rules holding in the data) with their scientific community. We review recent work on these topics, presenting general frameworks that we use to compare and contrast different approaches. We begin with the problem of focusing on different techniques of k-anonymity in section 2, we present and relate several important notions for this task, followed by k-anonymous patterns and frame work in section 3 and 4 respectively .We describe some general goals of different approaches and classify different techniques in section 5. In section 6, further relevant studies on these topic is illustrated and finally we end up with conclusion and future work in succeeding section.

II. PROBLEM DEFINITIONS

Let t be the initial micro data table and T be the released micro data table. T consists of a set of tuples over an attribute set. The attributes characterizing micro data are classified into the following three categories:

- Identifier attributes that can be used to identify a record such as Name and Medicare card.
- Quasi-identifier (QI) attributes that may be known by an intruder, such as Zip code and Age.

QI attributes are presented in the released micro data table T as well as in the initial micro data table t. Sensitive attributes that are assumed to be unknown to an intruder and need to be protected, such as Health Condition. Sensitive attributes are presented both in t and T .In what follows we assume that the identifier attributes have been removed and the quasi identifier and sensitive attributes are usually kept in the released and initial micro data table. Another assumption is that the value for the sensitive attributes is not available from any external source. This assumption guarantees that an intruder cannot use the sensitive attributes to increase the chances of disclosure. Unfortunately, an intruder may use record linkage techniques [1] between quasi-identifier attributes and external available information to glean the identity of individuals from the modified micro data.

To avoid this possibility of privacy disclosure, one frequently used solution is to modify the initial micro data; more specifically the quasi-identifier attributes values, in order to enforce the k-anonymity property.

Revised Manuscript Received on March 2012.

Swagatika Devi, Department Of Computer Science and Engineering, ITER Siksha 'O' Anusandhan University, Khandagiri Square, Bhubaneswar ,Orissa, India-751030 India, (Email: sweetsweettalk@gmail.com)

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Age	Country	Zip Code	Health Condition
<30	India	124**	Cancer
<30	India	124**	Cancer
<30	India	1242*	HIV
<30	India	1242*	HIV
>40	America	110**	Phthisis
>40	America	110**	Hepatitis
>40	America	110**	Heart Disease
>40	America	110**	Asthma
3*	India	1242*	Fever
3*	India	124**	Fever
3*	India	124**	Fever
3*	India	1242*	Indigestion

Table I Micro Data

Age	Country	Zip Code	Health Condition
<30	India	124**	Cancer
<30	India	124**	Cancer
<30	India	1242*	HIV
<30	India	1242*	HIV
>40	America	110**	Phthisis
>40	America	110**	Hepatitis
>40	America	110**	Heart Disease
>40	America	110**	Asthma
3*	India	1242*	Fever
3*	India	124**	Fever
3*	India	124**	Fever
3*	India	1242*	Indigestion

Table Ii A 2-Anonymous View Of Table I

Definition 1. (Quasi-identifier) A quasi-identifier (QI) is a minimal set Q of attributes in micro data table t that can be joined with external information to re-identify individual records (with sufficiently high probability).

Definition 2. (K-anonymity) The modified Micro data table T is said to satisfy k-anonymity if and only if each combination of quasi-identifier attributes in T occurs at least k times. A QI-group in the modified micro data T is the set of all records in the table containing identical values for the QI attributes. There is no consensus in the literature over the term used to denote a QI-group. This term was not defined when k-anonymity was introduced [3]. More recent papers use different terminologies such as equivalence class and QI-cluster [5]. For example, let the set {Age, Country, Zip Code} be the quasi-identifier of Table I. Table II is one 2-anonymous view of Table I since there are five QI-groups and the size of each QI-group is at least 2. So k- anonymity can ensure that even though an intruder knows a particular individual is in the k-anonymous micro data table t, s/he cannot infer which record in t corresponds to the individual with a probability greater than 1/k.

The k-anonymity property ensures protection against identity disclosure, i.e. the identification of an entity (person, institution). However, as we will show next, it does not protect the data against attribute disclosure, which occurs when the intruder finds something new about a target entity. Still consider the modified 2-anonymous table (Table II), where the set of quasi-identifier is composed of {Age, Country, and Zip Code} and Health Condition is the sensitive attribute. As we discussed above, identity disclosure does not happen in this modified micro data. However, assuming that external information in Table III is available, attribute disclosure can take place. If the intruder knows that in the modified table (Table II) the Age attribute was modified to '<30', he can deduce that both Rick and Rudy have Cancer, even he does not know which record, 3 or 4, corresponds to which person. This example shows that even if k-anonymity can well protect identity disclosure, sometimes it fails to protect against sensitive attribute disclosure. A similar privacy model, called l-diversity, is in [6].

Name	Age	Country	Zip Code
Ram	26	UP	12426
Haseen	45	India	11064
Manas	25	USA	12429
Asis	48	China	11074

Table III External Available Information

Category ID	Sensitive attribute vales	Sensitivity
One	HIV,Cancer	Top Secret
Two	Phthisis, Hepatitis	Secret
Three	Heart Disease, Asthma	Less Secret
Four	Fever, Indigestion	Non Secret

Table IV Categories Of Health Condition

Definition 3. (P-sensitive k-anonymity) The modified micro data table T satisfies p-sensitive k- anonymity property if it satisfies k-anonymity, and for each QI-group in T, the number of distinct values for each sensitive attribute occurs at least p times within the same QI-group. Although the p-sensitive k-anonymity represents an important step beyond k-anonymity in protecting against attribute disclosure, it still has some shortcomings. Following through, we show that p-sensitive k-anonymity is insufficient to prevent Similarity Attack. Similarity Attack: When the sensitive attribute values in a QI-group are distinct but similar sensitivity, an adversary can learn important information. Sometimes, the domain of the sensitive attributes, especially the categorical ones, can be partitioned into categories according to the sensitivity of attributes. For example, in medical data sets Table I, the Health Condition attribute can be classified into four categories (see Table IV).

The different types of diseases are organized in a category domain. The attribute values are very specific, for example they can represent HIV or Cancer, which are both Top Secret information of the individuals.

ID	Age	Country	Zip Code	Health Condition	Category ID
1	<40	India	1242*	HIV	One
4	<40	India	1242*	Cancer	One
9	<40	India	1242*	Fever	Four
12	<40	India	1242*	Indigestion	Four
5	>40	America	110**	Hepatitis	Two
6	>40	America	110**	Phthisis	Two
7	>40	America	110**	Asthma	Three
8	>40	America	110**	Heart Disease	Three
2	<40	India	1240*	Cancer	One
3	<40	India	1240*	HIV	One
10	<40	India	1240*	Fever	Four
11	<40	India	1240*	Fever	Four

Tablev 2-Sensitive4-Anonymous Micro Data

In the case that the initial micro data contains specific sensitive attributes like Health Condition, the data owner can be interested in protecting not only these most specific values, but also the category that the sensitive values belong to. For example, the information of a person affected with Top Secret needs to be protected, no matter whether it is HIV or Cancer.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

If we modify the micro data to satisfy p-sensitive k-anonymity property, it is possible that in a QI-group with p distinct sensitive attribute values, all of them belong to the same pre-defined category. For instance, the values {HIV, HIV, Cancer, and Cancer} in one QI-group in Table V all belong to Top Secret category. To avoid such situations, we introduce our new enhanced p-sensitive k-anonymity models, which are aware of not only protecting specific sensitive values.

III. K-ANONYMOUS PATTERNS

We start by defining binary databases and patterns following the notation in [12].

Definition 1. A binary database D = (I, T) consists of a finite set of binary variables I = {i1…….ip}, also known as items, and a finite multiset T = { t1…….. tn } of p-dimensional binary vectors recording the values of the items. Such vectors are also known as transactions. A pattern for the variables in I is a logical (propositional) sentence built by AND (∩), OR (U) and NOT (¬) logical connectives, on variables in I. The domain of all possible patterns is denoted Pat (I). One of the most important properties of a pattern is its frequency in the database, i.e. the number of individuals (transactions) in the database which make the given pattern true.

Definition 2. Given a database D, a transaction t ε D and a pattern p, we write p(t) if t makes p true. The support of p in D is given by the number of transactions which makes p true: $sup_D(p) = I\{t \varepsilon D \; I \; p(t)\}I$.

The most studied pattern class is the item set, i.e., a conjunction of positive valued variables, or in other words, a set of items. The retrieval of item sets which satisfy a minimum frequency property is the basic step of many data mining tasks.

Definition 3. The set of all item sets 2^I, is a pattern class consisting of all possible conjunctions of the form i1 ^i2 ^:….^ im . Given a database D and a minimum support threshold σ , the set of σ -frequent item sets in D is denoted $F(D; \sigma) = \{<X, sup_D(X)>IX \varepsilon 2^I \cap sup_D(X) \geq \sigma \}$ Item sets are usually denoted in the form of set of the items in the conjunction, e.g. $\{i_1…..im\}$; or sometimes, simply i1 ….. im. Figure 1 shows the different notation used for general patterns and for item sets.

Definition 4. Given a database D and an anonymity threshold k, a pattern p is said to be k-anonymous if $sup_D(p) \geq k$ or $sup_D(p) = 0$.

IV. THE K-ANONYMITY FRAMEWORK

In many applications, the data records are made available by simply removing key identifiers such as the name and social-security numbers from personal records. However, other kinds of attributes (known as pseudo-identifiers) can be used in order to accurately identify the records.

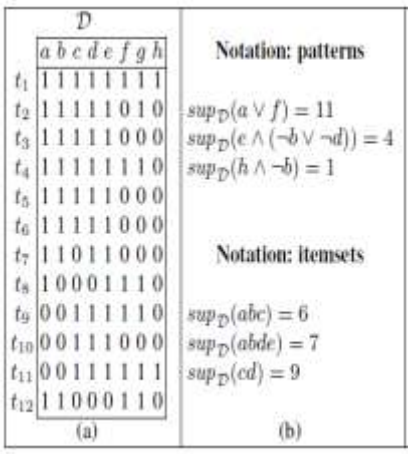

Fig.1 (a) Binary Database (b) Notations for patterns and item set

For example, attributes such as age, zip-code and sex are available in public records such as census rolls. When these attributes are also available in a given data set, they can be used to infer the identity of the corresponding individual. A combination of these attributes can be very powerful, since they can be used to narrow down the possibilities to a small number of individuals. In k-anonymity techniques [8], we reduce the granularity of representation of these pseudo-identifiers with the use of techniques such as generalization and suppression. In the method of generalization, the attribute values are generalized to a range in order to reduce the granularity of representation. For example, the date of birth could be generalized to a range such as year of birth, so as to reduce the risk of identification. In the method of suppression, the value of the attribute is removed completely. It is clear that such methods reduce the risk of identification with the use of public records, while reducing the accuracy of applications on the transformed data. In order to reduce the risk of identification, the k-anonymity approach requires that every tuple in the table be indistinguishability related to no fewer than k respondents. This can be formalized as follows:

Definition 1. Each release of the data must be such that every combination of values of quasi-identifiers can be indistinguishably matched to at least k respondents.

We note that the problem of optimal anonymization is NP-hard. Nevertheless, the problem can be solved quite effectively by the use of a number of heuristic methods. A method is the k-Optimize algorithm which can often obtain effective solutions. The approach assumes an ordering among the quasi-identifier attributes. The values of the attributes are discretized into intervals (quantitative attributes) or grouped into different sets of values (categorical attributes). Each such grouping is an item. For a given attribute, the corresponding items are also ordered. An index is created using these attribute-interval pairs (or items) and a set enumeration tree is constructed on these attribute-interval pairs. This set enumeration tree is a systematic enumeration of all possible generalizations with the use of these groupings.

The root of the node is the null node, and every successive level of the tree is constructed by appending one item which is lexicographically larger than all the items at that node of the tree. We note that the number of possible nodes in the tree increases exponentially with the data dimensionality. Therefore, it is not possible to build the entire tree even for modest values of n. However, the k-Optimize algorithm can use a number of pruning strategies to good effect. In particular, a node of the tree can be pruned when it is determined that no descendent of it could be optimal. This can be done by computing a bound on the quality of all descendents of that node, and comparing it to the quality of the current best solution obtained during the traversal process. A branch and bound technique can be used to successively improve the quality of the solution during the traversal process. In general, the Incognito algorithm computes (i + 1)-dimensional generalization candidates from the i-dimensional generalizations, and removes all those generalizations which do not satisfy the k-anonymity constraint. This approach is continued until, no further candidates can be constructed, or all possible dimensions have been exhausted. We note that generalization and suppression are not the only transformation techniques for implementing k-anonymity. Micro-aggregation in which clusters of records are constructed can also be used. For each cluster, its representative value is the average value along each dimension in the cluster to design the clustering. In [8], a related method has been independently proposed for condensation based privacy-preserving data mining. This technique generates pseudo-data from clustered groups of k-records. The process of pseudo-data generation uses principal component analysis of the behavior of the records within a group. Another technique proposed in [11] uses genetic algorithms in order to construct k anonymous representations of the data. Both of these techniques require high computational times, and provide no guarantees on the quality of the solutions found. The only known techniques which provide guarantees on the quality of the solution are approximation algorithms ([9], [14]), in which the solution found is guaranteed to be within a certain factor of the cost of the optimal solution.

V. CLASSIFICATION OF K-ANONYMITY TECHNIQUES

The original k-anonymity proposal considers the application of generalization at the attribute (column) level and suppression at the tuple (row) level. However, both generalization and suppression can also be applied, and have been investigated, at a finer granularity level. We discuss the different ways in which generalization and suppression can be applied.

Generalization can be applied at the level of:

Attribute (AG): generalization is performed at the level of column; a generalization step generalizes all the values in the column

	Suppression			
Generalization	*Tuple*	*Attribute*	*Cell*	*None*
Attribute	AG_TS	AG_AS ≡ AG_	AG_CS	AG_ ≡ AG_AS
Cell	CG_TS not applicable	CG_AS not applicable	CG_CS ≡ CG_	CG_ ≡ CG_CS
None	_TS	_AS	_CS	- not interesting

Fig. 2 Classification of K-Anonymity techniques

Cell (CG): generalization is performed on single cells; as a result a generalized table may contain, for a specific column, values at different generalization levels. For instance, in the Date of birth column some cells can report the specific day (no generalization), others the month (one step of generalization), others the year (two steps of generalization), and so on. Generalizing at the cell level has the advantage of allowing the release of more specific values (as generalization can be confined to specific cells rather than hitting whole columns). However, besides a higher complexity of the problem, a possible drawback in the application of generalization at the cell level is the complication arising from the management of values at different generalization levels within the same column.

Suppression can be applied at the level of:

Tuple (TS): suppression is performed at the level of row; a suppression operation removes a whole tuple.

Attribute (AS): suppression is performed at the level of column, a suppression operation obscures all the values of a column.

Cell (CS): suppression is performed at the level of single cells; as a result a k-anonymized table may wipe out only certain cells of a given tuple/attribute.

Below we discuss the different models resulting from our classification, characterize them, and classify existing approaches accordingly. We refer to each model with a pair (separated by _), where the first element describes the level of generalization (AG, CG, or none) and the second element describes the level of suppression(TS,AS, CS, or none). Table in Fig. 2 summarizes these models.

AG _TS Generalization is applied at the level of attribute (column) and suppression at the level of tuple (row). It enjoys a trade of between the computational complexity and the quality of the anonymized table.

AG_ AS Both generalization and suppression are applied at the level of column. No specific approach has investigated this model. It must also be noted that if attribute generalization is applied, attribute suppression is not needed; since suppressing an attribute (i.e., not releasing any of its values) to reach k-anonymity can equivalently be modeled via a generalization of all the attribute values to the maximal element in the value hierarchy. This model is then equivalent to model AG (attribute generalization, no suppression).

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

AG _CS Generalization is applied at the level of column, while suppression at the level of cell. It allows to reduce the effect of suppression, at the price however of a higher complexity of the problem.

AG_ Generalization is applied at the level of column, suppression is not considered. As noted above, it is equivalent to model AG_ AS.

CG _CS Both generalization and suppression are applied at the cell level. Then, for a given attribute we can have values at different levels of generalization. By observations similar to those illustrated for AG _AS, this model is equivalent to CG (cell generalization, no suppression). Indeed, suppression of a cell can be equivalently modeled as the generalization of the cell at the maximal element of the value hierarchy.

CG_ Generalization is applied at the level of cell, suppression is not considered. As just noted, it is equivalent to CG _CS.

TS Suppression is applied at the tuple level, generalization is not allowed. No approach has investigated this model, which however can be modeled as a reduction of AG TS to the case where all the generalization hierarchies have height zero (i.e., no hierarchy is defined).

AS Suppression is applied at the attribute level, generalization is not allowed. No explicit approach has investigated this model. We note, however, that it can be modeled as a reduction of AG where all the generalization hierarchies have height of 1.

CS Suppression is applied at the cell level, generalization is not allowed . Again, it can be modeled as a reduction of AG where all the generalization hierarchies have height of 1. In addition to these models, we have the obvious uninteresting combination (no generalization, no suppression) and two models, which are not applicable, namely: CG _TS (cell generalization, tuple suppression) and CG _AS (cell generalization, attribute suppression). The reason for their non applicability is that since generalizing a value at the maximum element in the value hierarchy is equivalent to suppressing it, supporting generalization at the fine grain of cell clearly implies the ability of enforcing suppression at that level too.

Note that, because of the equivalence relationships pointed out in the discussion above, there are essentially seven possible models.

VI. FURTHER STUDIES ON K-ANONYMITY

We now briefly survey some interesting studies based on the concept of k- anonymity.

A.K-Anonymity for Protecting Location Privacy

The k-anonymity property has been studied also for protecting location privacy. In the context of location-based services, Bettini, Wang and Jajodia [6] present a framework for evaluating the privacy of a user identity when location information is released. In this case, k-anonymity is guaranteed, not among a set of tuples of a database, but in a set of individuals that can send a message in the same spatio-temporal context.

B. Distributed Algorithms for K-Anonymity

In many cases, it is important to maintain k anonymity across different distributed parties. The broad idea is for the two parties to agree on the quasi-identifier to generalize to the same value before release. A similar approach is discussed in [15], in which the two parties agree on how the generalization is to be performed before release. In [16], an approach has been discussed for the case of horizontally partitioned data. The work in [16] discusses an extreme case in which each site is a customer which owns exactly one tuple from the data. It is assumed that the data record has both sensitive attributes and quasi-identifier attributes. The solution uses encryption on the sensitive attributes. The sensitive values can be decrypted only if therefore are at least k records with the same values on the quasi-identifiers. Thus, k-anonymity is maintained. The issue of k anonymity is also important in the context of hiding identification in the context of distributed location based services . In this case, k-anonymity of the user-identity is maintained even when the location information is released. Such location information is often released when a user may send a message at any point from a given location. A similar issue arises in the context of communication protocols in which the anonymity of senders (or receivers) may need to be protected. A message is said to be sender k-anonymous, if it is guaranteed that an attacker can at most narrow down the identity of the sender to k individuals. Similarly, a message is said to be receiver k-anonymous, if it is guaranteed that an attacker can at most narrow down the identity of the receiver to k individuals.

C.K-Anonymity for Communication Protocols

K-anonymity has also been investigated to preserve privacy in communication protocols ([7],[14]) with the notion of sender (receiver, resp.) k-anonymity. A communication protocol is sender k-anonymous (receiver k-anonymous, resp.) if it guarantees that an attacker, who is trying to discover the sender (receiver, resp.) of a message, can just detect a set of k possible senders (receivers, resp.).

VII. CONCLUSIONS

A key issue in measuring the security of different privacy-preservation methods is the way in which the underlying privacy is quantified. The idea in privacy quantification is to measure the risk of disclosure for a given level of perturbation. The k-anonymity is an attractive technique because of the simplicity of the definition and the numerous algorithms available to perform the anonymization. Nevertheless the technique is susceptible to many kinds of attacks especially when background knowledge is available to the attacker. Clearly, while k-anonymity is effective in preventing identification of a record, it may not always be effective in preventing inference of the sensitive values of the attributes of that record. Therefore, the technique of l-diversity was proposed which not only maintains the minimum group size of k, but also

focuses on maintaining the diversity of the sensitive attributes. When the number of attributes in the quasi-identifier increases, the information loss of the resulting k-anonymized table may become very high.

The intuition behind this result is that the probability that k tuples in the private table are similar" (i.e., they correspond to the same tuple in the anonymized table with a reduced loss of information) is very low. The ability to identify minimal quasi-identifiers is therefore important.

REFERENCES

1. W. E. Winkler. Advanced Methods for Record Linkage, Proceedings of the Section on Survey Research Methods, American Statistical Society, 467-472.
2. R. Agrawal and R. Srikant. Privacy-preserving data mining. In Proceedings of the 2000 ACM SIGMOD on Management of Data.
3. P. Samarati. Protecting respondents' identities in micro data release. IEEE Transactions on Knowledge and Data Engineering, 13(6):1010-1027. 2001.
4. V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, and Y. Theodoridis. State-of-the-art in privacy preserving data mining. SIGMOD Rec., 33(1):50.57, 2004.
5. T. M. Truta, A. Campan and P. Meyer. Generating Micro data with p-sensitive k-anonymity Property. SDM 2007: 124-141
6. A.Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam. l-Diversity: Privacy beyond k-anonymity. In ICDE, 2006.
7. D. Agrawal and C. C. Aggarwal. On the design and quantification of privacy preserving data mining algorithms. In Proceedings of the twentieth ACM PODS, 2001.
8. P. Samarati: Protecting Respondents' Identities in Microdata Release. IEEE Trans. Knowl. Data Eng. 13(6): 1010-1027 (2001).
9. G. Aggarwal, T. Feder ,K. Kenthapadi,R. Motwani, R. Panigrahy, D. Thomas , A. Zhu : Anonymizing Tables. ICDT Conference, 2005.
10. C. Bettini , Wang XS, S. Jajodia (2005). Protecting privacy against location- based personal identification. In Proc. of the Secure Data Management, Trondheim, Norway.
11. V.S. Iyengar : Transforming Data to Satisfy Privacy Constraints. KDD Conference, 2002.
12. D. Hand, H. Mannila, and P. Smyh. Principles of Data Mining. The MIT Press, 2001.
13. M. Kantarcioglu, J. Jin, and C. Clifton. When do data mining results violate privacy? In Proceedings of the tenth ACM SIGKDD, 2004.
14. G. Aggarwal, T.Feder ,K. Kenthapadi,R. Motwan, R. Panigrahy, D. Thomas ,A. Zhu: Approximation Algorithms for k-anonymity. Journal of Privacy Technology, paper 20051120001, 2005.
15. K. Wang , B.C.M. Fung , G.Dong : Integarting Private Databases for Data Analysis. Lecture Notes in Computer Science, 3495, 2005.
16. S. Zhong S., Z. Yang , R. Wright : Privacy-enhancing k-anonymization of customer data, In Proceedings of the ACM SIGMOD-SIGACT-SIGART Principles of Database Systems, Baltimore, MD. 2005.

AUTHOR PROFILE

SWAGATIKA DEVI received her B.Tech and M.Tech degrees in Computer Science and Engineering. She is working as an assistance professor in the department of Computer Science and Engg.in SOA University, Bhubaneswar, Orissa.She got gold medal during her M.Tech career. She contributed many international journals and presented research papers in different international conferences and also has tremendous contribution towards research book chapters. Her research interests include soft computing.,data mining and bio-informatics.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication