International Journal of Research
Available at https://edupediapublications.org/journals

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 09
August 2017

Confidentiality Based Data Transfer Using Privacy Preserving Association Rule Mining

V.Neha & K. David Raju

[1](M.Tech CSE Pursuing) (St.Peters Engineering college, Hyderabad, TS, INDIA)

[2]Assoc. Professor, Dept.of CSE) (St.Peters Engineering college, Hyderabad, TS, INDIA)

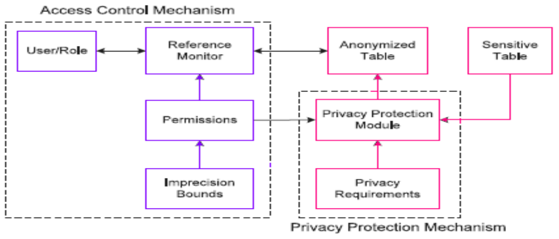
neharao0611@gmail.com & davidraju@stpetershyd.com

Abstract:

Privacy Preserving Association Rule Mining (PPAM) becomes an important issue in recent years.Since data mining alone is not enough to share data between companies without privacy preserving. In this paper, a new technique has been proposed to maintain the confidentiality of the data by using Privacy without returning to mining sensitive data again. And also been achieved high speed and fewer memory requirements.

INTRODUCTION

Many organizations gather, do some research on consumer information for improving facilities of them. ACMs (Access Control Mechanisms) are used for making sure that the valid data are accessible to user. However, there may be a chance that the critical data may be ill-used by some of official users for customers secrecy. And the idea of secrecy maintenance for critical information needs the enforcement of policies related to privacy or the protection of identity expose by meeting privacy needs. Here, we do research secrecy maintenance from ambiguity phase. After the elimination of recognizing attributes on delicate data, even is still vulnerable for connecting assaults by the official users. And this drawback is researched widely at micro information publishing zones. and secrecy related explanations, example, l-diversity , k-anonymity , and variance variety. The Ambiguity algorithm does use generalization, records' suppression for satisfying security needs with the help of smallest spin of micro information. In order to go through privacy and security of critical data, we use ambiguity techniques with an entrée controlling scheme. We can reach secrecy at the cost of accurateness, under an access control policy, the roughness(imprecision) is presented in official data.The idea of imprecision limits are used for all agreements for definingthe threshold on imprecision amount that is accepted. The current workload known ambiguity methods minimize the aggregate of imprecision for every query, the imprecision that is appended to every consent or query in the anonymized micro information is not seen. Making the secrecy need more strict (example., increasing the k value or l) outcomes in further imprecision for queries. However, the fault of satisfying accuracy restraints for individual permissions in a rule or workload is not reviewed previously.The algorithms that are suggested in this article (heuristics) are related to situation of workload concerned ambiguity. The ambiguity to constant data publication is viewed in the literature survey section. Here, in the, article, the main attention is on stationary table. To represent our method, we assume a role-based entree controller. However, idea of correctness restraints to permissions can be useful for all secrecy maintenance strategy, example, unrestricted entree control.

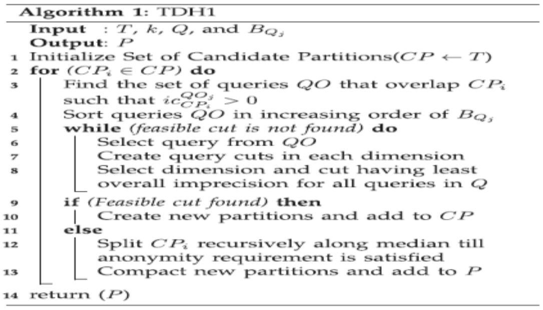


SYSTEM ANALYSIS

Different ORGANIZATIONS gather and do study the customer data for improving services of them. ACMs (stands for Access Control Mechanisms) are utilized for making sure that only official data is accessible to the user. However, delicate data may be still distorted by official user for compromising the consumer secrecy. The privacy-preservation mechanism for delicate information can need the application of secrecy policy or the defense over identity theft by fulfilling secrecy needs.Existing workload related ambiguity methodologies do lower the imprecision sum to every query, the imprecision that is increased to each permission or query in the anonymised micro information isn't recognized. Building the privacy need much severe (e.g., increasing the value of k or l) results in additional imprecision for queries.

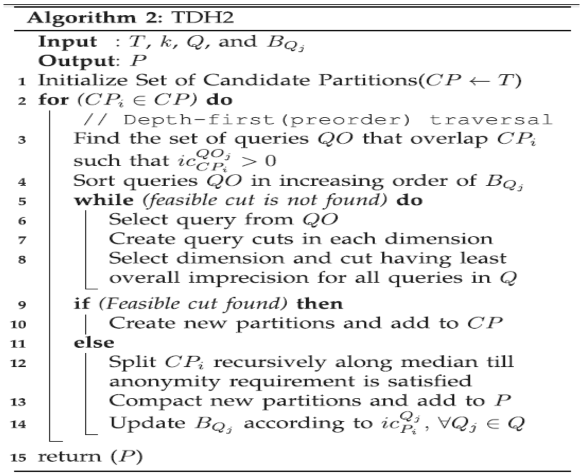
Top-Down Heuristic:

In top down Heuristic, plits are done along the median. Assume if a query is overlapped by a partition, and id the median is also in query, then after partitioning, the query imprecision will not change as query is overlapped by both the partitions as illustrated . here, we proposed for splitting the partitions along the query cut and then select the dimensions by which the imprecisions is low for every query. If a partition is overlapped by multiple queries, then the query that is to be utilized for the cut is chosen. The queries with >0 imprecision for a particular partitions are arranged on the basis of the imprecision limit and the query that is having lower imprecisions bounds is chosen. The intuition behind this decision is that the queries with smaller bounds have lower tolerance for error and such a partition split ensures the decrease in imprecision for the query with the smallest imprecision bound. All feasible cuts failed to satisfy the secrecy requirements, then the next query in the arranged set is utilized for checking for partition's partition.



```
Algorithm 1: TDH1
   Input  : $T$, $k$, $Q$, and $B_{Q_j}$
   Output: $P$
 1 Initialize Set of Candidate Partitions($CP \leftarrow T$)
 2 for ($CP_i \in CP$) do
 3     Find the set of queries $QO$ that overlap $CP_i$
       such that $ic_{CP_i}^{QO_j} > 0$
 4     Sort queries $QO$ in increasing order of $B_{Q_j}$
 5     while (feasible cut is not found) do
 6         Select query from $QO$
 7         Create query cuts in each dimension
 8         Select dimension and cut having least
           overall imprecision for all queries in $Q$
 9     if (Feasible cut found) then
10         Create new partitions and add to $CP$
11     else
12         Split $CP_i$ recursively along median till
           anonymity requirement is satisfied
13         Compact new partitions and add to $P$
14 return ($P$)
```

TDH2 In TDH2, As splits are included in outcome, the query bounds are updated. This updation is carried out with the subtraction of $ic_{P_i}^{Q_j}$ value from B_{Q_j}, (which is the imprecision bound of all queries), for a split (partition) Pi, that is included to outcome. E.g. , consider a partition with its size as k, and it has the imprecisions 5, 10 for the Queries Q1, Q2 with the imprecision bounds 100, 200, then limits will be exchanged to 95, 190. If the kd-tree traversal is a dept first(preorder), then we can achieve the best results. This preorder traversal makes ensure that the specified partition is recursively partitioned until final node(leaf node) can be found. Following diagram that is there belove, lists TDH2 algorithm. If we comare TDH1 algorith with TDH2 algorithm, we can find two differences. First one is, from line2 to line 14, the kd-tree loop is preorder. Second one is,if the partitions are added to output (P), then the query bounds are updated (in Line 14). For both the TDH1 and TDH2 algorithms, the time complexity is same, which is $\mathcal{O}(d|Q|^2n^2)$,

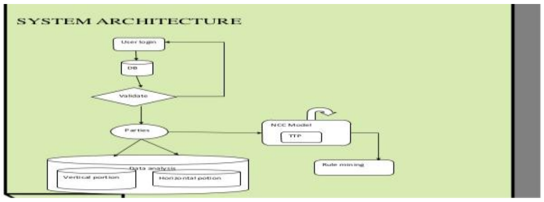


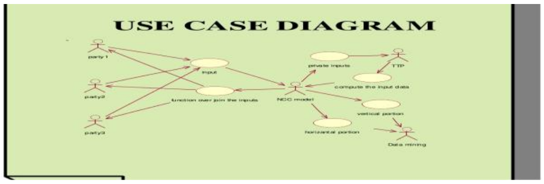
```
Algorithm 2: TDH2
   Input  : $T$, $k$, $Q$, and $B_{Q_j}$
   Output: $P$
 1 Initialize Set of Candidate Partitions($CP \leftarrow T$)
 2 for ($CP_i \in CP$) do
        // Depth-first(preorder) traversal
 3     Find the set of queries $QO$ that overlap $CP_i$
       such that $ic_{CP_i}^{QO_j} > 0$
 4     Sort queries $QO$ in increasing order of $B_{Q_j}$
 5     while (feasible cut is not found) do
 6         Select query from $QO$
 7         Create query cuts in each dimension
 8         Select dimension and cut having least
           overall imprecision for all queries in $Q$
 9     if (Feasible cut found) then
10         Create new partitions and add to $CP$
11     else
12         Split $CP_i$ recursively along median till
           anonymity requirement is satisfied
13         Compact new partitions and add to $P$
14         Update $B_{Q_j}$ according to $ic_{P_i}^{Q_j}$, $\forall Q_j \in Q$
15 return ($P$)
```

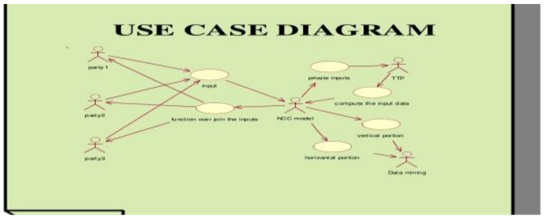
TDH3 In the TDH3 algorithm, TDH2 algorithm is modified so that the time complexity of $\mathcal{O}(d|Q|nlgn)$ may be gotten at reduced precision cost in query outcomes. For given split, for a query with a lowest imprecision bound, the TDH3 algorithm checks the query cuts. And also, the second one is the querycuts are realistic if ration of size of the obtaining partition is highly straight. For this algorithm (TDH3), a threshold we use 1:99 skew ratio. If the result of query cut of a partition is with size bigger than 100 time the others, we can ignore tha. TDH3 algorithm represented below in Algorithm 3. In 4th line of this algorithm, only one query is used as the candidate cut. Coming to line number 6, for a feasible cut the ratio of partition size to be satisfied. If it couldn’t find a feasible query cut, then the partition is split along the median (as in Line 11).

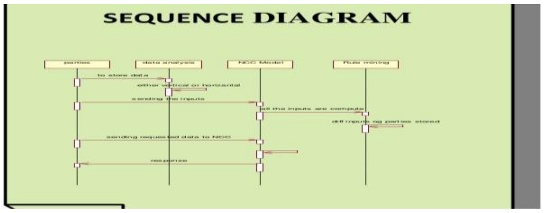
```
Algorithm 3: TDH3
   Input  : $T$, $k$, $Q$, and $B_{Q_j}$
   Output: $P$
1  Initialize Set of Candidate Partitions($CP \leftarrow T$)
2  for ($CP_i \in CP$) do
       // Depth-first(preorder) traversal
3      Find the set of queries $QO$ that overlap $CP_i$
       such that $ic_{CP_i}^{QO_j} > 0$
4      Select query from $QO$ with smallest $B_{Q_j}$
5      Create query cuts in each dimension
6      Reject cuts with skewed partitions
7      Select dimension and cut having least overall
       imprecision for all queries in $Q$
8      if (Feasible cut found) then
9          Create new partitions and add to $CP$
10     else
11         Split $CP_i$ recursively along median till
           anonymity requirement is satisfied
12         Compact new partitions and add to $P$
13         Update $B_{Q_j}$ according to $ic_{P_i}^{Q_j}$, $\forall Q_j \in Q$
14 return ($P$)
```

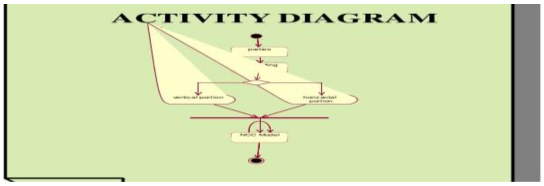
Extension In this extended model, a user query is modified by the access control mechanism and only the authorized tuples are returned. Since Cell level access control is proving a best solution than column level access, along with Column level access we also implemented cell level access control and incremental data for our system.If given a relation **T = {A$_1$,A$_2$, . . .,A$_n$, . . A$_{2n}$}**, where Ai is an attribute, T* is the anonym zed version of the relation T. We assume that T is a static relational table. Her on this relation T, we performed cell level access control over it. Cell level access control for relational data is implemented by replacing the unauthorized cell values by NULL values.RBAC (stands for Role-based Access Control) allows defining approvals on object on the basis of role in an corporation. An RBAC rule configuration consists of group of Users (U), and a group of Roles (R),a group of Permissions (P). For the relational RBAC model, we assume that the selection predicates on the QI attributes define a permission. UA is a user to role (U X R) task relative, PA is role to permission (R X P) task relation.A role hierarchy (RH) defines an inheritance relationship among roles and is a partial order on roles (R X R). Each permission defines a hyper-rectangle in the tuple space and all the tuples enclosed by this hyper-rectangle are authorized to the role assigned to the permission. In practice, when a user assigned to a role executes a query, the tuples satisfying the conjunction of the query predicate and the permission are returned

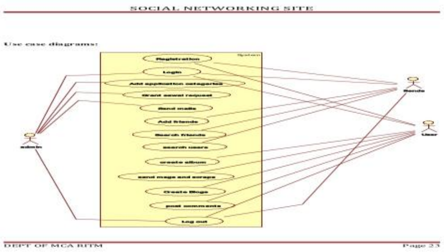

SYSTEM ARCHITECTURE


USE CASE DIAGRAM


USE CASE DIAGRAM

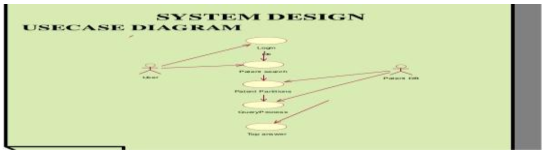

SEQUENCE DIAGRAM
parties
data analysis
NCC Model
Rule mining
to store data
either vertical or horizontal
sending the inputs
all the inputs are compute
diff inputs og parties stored
sending requested data to NCC
response

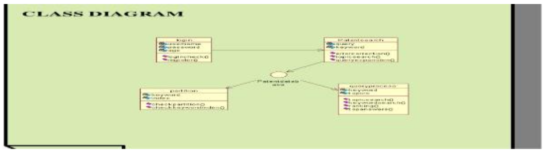

ACTIVITY DIAGRAM

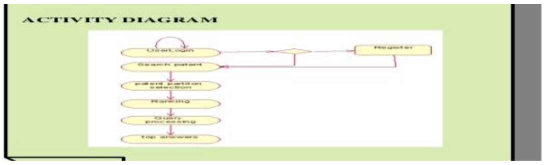

SOCIAL NETWORKING SITE

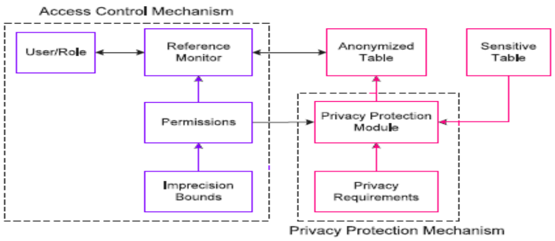
Use case diagrams:

DEPT OF MCA BITM   Page 23









SYSTEM TESTING

Main aim of system test is for finding out mistakes. It is a manner of attempting for uncovering each fault or weakness of our work product. The Testing offers an approach for checking the functionality of the components of our project, sub assemblages, assemblages and / or a completed outcome. And is a manner of training software with a commitment of making sure that the scheme encounters it's needs, user expectations, doesn't fail in an offensive process. We have different sorts of tests. Every type of test focuses on particular testing needs

Unit testing

It involves the testing of internal design and to test the designed program logic is functioning properly or not. The code that we write (code flow and decision branches) must be validated in this testing phase only. And we can also say that this is examining of separate components of our system, and is completed after finishing the component before assembling. This structural test trussts on wisdom of it's building and is offensive.

Integration testing

Integrated software components are being tested by Integration testing for determining weather they usually run as a single package. This test event that is focused is much related to the fundamental output of fields/screens. The Integration testing express that even though the units are satisfied separately, as revealed with the help of unit testing, the component combination is consistent and correct. This Integration testing is particularly focused at revealing the problems that rise from the component combination.

Functional test

The Functional testing offers regular demos that tasks tested are accessible as stated by the requirements of technical and business needs, user manual and the system documentation. Functional test is focused on below elements: The Organizing and planning of functionals testing is driven on needs, other exclusive testing cases and key functions. In add-on, the systematic exposure related for identifying the flow of predefined processes, data fields, Business process, and consecutive progress should be taken to test. Prior to finishing the functionals test, some other test cases are recognized, the current test's active worth is defined.

System Testing

The System test makes sure that whole united software scheme encounters needs. It does test a design for making sure well-identified and probable outcomes. The configuration oriented system integration test is an instance of system testing. And this System testing is based on procedure descriptions and does flow, stressing integration points and pre-driven process links.

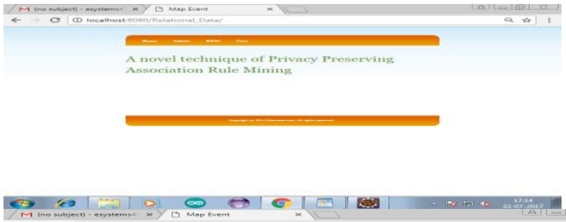
White Box Test

It is a test case, in that the software testers have awareness of internal working, language, software structure, or at least intention of it. It is utilized for examine the zones that cannot be touched from the black box testing level

Black Box Test

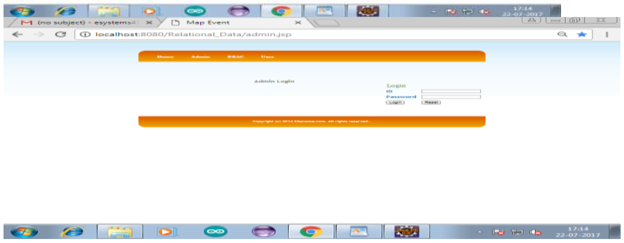
In this test case, testing of the software is done with no awareness of internal functionalities of a system, language or structure of modules that is tested. The Black box testing cases are the utmost another sort of test, and these should be penned from a standard document of source, such as requirements/specification documents. In this test, the software under test is discussed as a black box. The test does provide inputs, answers to outcomes without seeing how software does work.

SCREEN SHOTS


(no subject) - esystems4 × Map Event ×
localhost:8080/Relational_Data/

A novel technique of Privacy Preserving Association Rule Mining

17:14
22-07-2017

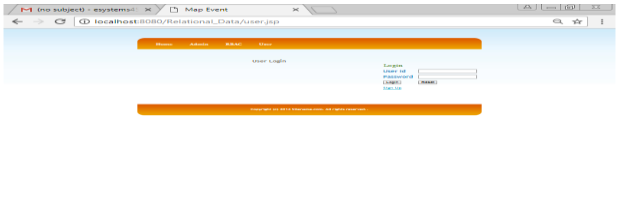

(no subject) - esystems4 × Map Event ×
localhost:8080/Relational_Data/admin.jsp

Admin Login

Login
ID
Password
Login Reset

Copyright (c) 2014 Sitename.com. All rights reserved.

17:14
22-07-2017

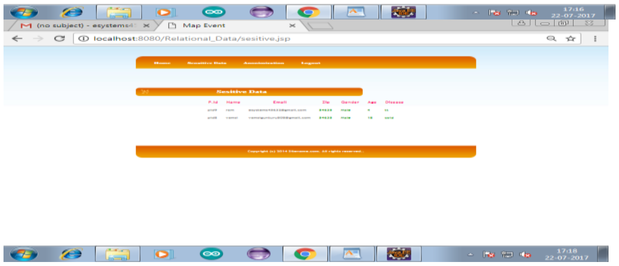

(no subject) - esystems4 × Map Event ×
localhost:8080/Relational_Data/user.jsp

User Login

Login
User Id
Password
Login Reset
Sign Up

Copyright (c) 2014 Sitename.com. All rights reserved.

17:16
22-07-2017


(no subject) - esystems4 × Map Event ×
localhost:8080/Relational_Data/sesitive.jsp

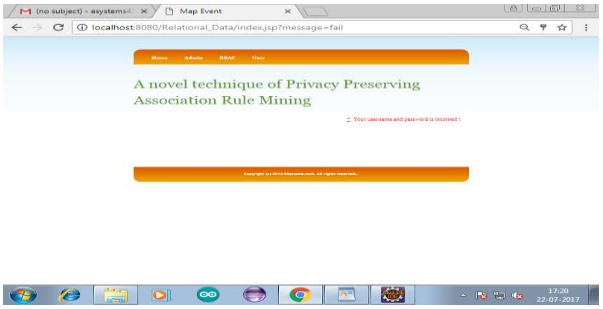
Home Sensitive Data Anonimization Logout

Sesitive Data

P.Id	Name	Email	Zip	Gender	Age	Disease
[illegible]	ram	[illegible]@gmail.com	[illegible]	Male	[illegible]	[illegible]
[illegible]	vamsi	[illegible]@gmail.com	[illegible]	Male	[illegible]	[illegible]

Copyright (c) 2014 Sitename.com. All rights reserved.

17:18
22-07-2017


(no subject) - esystems4 × Map Event ×
localhost:8080/Relational_Data/index.jsp?message=fail

A novel technique of Privacy Preserving Association Rule Mining

Your username and password is incorrect !

17:20
22-07-2017

Conclusion

Since a previous works need a further study about time efficient, data size, data form, truly hiding association rule. This paper proposes a novel technique to overcome the problems of PPDM and PPAM techniques in a specialist. The proposed technique uses a privacy association rules directly without a return to mining again of an original sensitive dataset. So this technique has good properties in a time efficient. Also, it can try to any size and type of data. It allows the individual select their security level in easy.

REFERENCES

[1] M. N. Dehkordi, K. Badie, and A. K. Zadeh, “A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms”, Journal of software, VOL. 4, NO. 6, pp. 555-562, August 2009.

[2] D. Thakur, and H. Gupta, “An Exemplary study of privacy preserving association rule mining techniques”, IJARCSSE, Vol. 3, Issue 11,pp. 893-900,November 2013.

[3] K. Sathiyapriya1 and G. S. Sadasivam, “A Survey on privacy preserving association rule mining”, IJDKP, Vol.3, No.2, pp. 191-131, March 2013.

[4] K. Saranya, K. Premalatha, and S. Rajasekar, “A survey on privacy preserving data mining”, IEEE, Electronics and Communication Systems (ICECS), 2015 2nd International Conference on, pp.1740 – 1744.

[5] V. Garg, A. Singh, and D. Singh, ”A Survey of Association Rule Hiding Algorithms”,IEEE,

Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on, pp. 404 – 407.

[6] S. B. Sadkhan and R. S. Mohammed, "Proposed random unified chaotic map as PRBG for voice encryption in wireless communication", International Conference on Communication, Management, and Information Technology "ICCMIT'15" April 20-22, 2015 Prague-Czech.

[7] BenSa¨ida, "A practical test for noisy chaotic dynamics", ELSEVIER, 2015.

[8] K. Browder and M. Davidson, "The Virtual Private Database in oracle9ir2," Oracle Technical White Paper, vol. 500, 2002.

[9] A. Rask, D. Rubin, and B. Neumann, "Implementing Row-and Cell-Level Security in Classified Databases Using SQL Server 2005," MS SQL Server Technical Center, 2005.

[10] S. Rizvi, A. Mendelzon, S. Sudarshan, and P. Roy, "Extending Query Rewriting Techniques for Fine-Grained Access Control," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 551-562, 2004.

[11] S. Chaudhuri, T. Dutta, and S. Sudarshan, "Fine Grained Authorization through Predicated Grants," Proc. IEEE 23rd Int'l Conf. Data Eng., pp. 1174-1183, 2007.

[12] K. LeFevre, R. Agrawal, V. Ercegovac, R. Ramakrishnan, Y. Xu, and D. DeWitt, "Limiting Disclosure in Hippocratic Databases," Proc. 30th Int'l Conf. Very Large Data Bases, pp. 108-119, 2004.

[13] D. Ferraiolo, R. Sandhu, S. Gavrila, D. Kuhn, and R. Chandramouli, "Proposed NIST Standard for Role-Based Access Control," ACM Trans. Information and System Security, vol. 4, no. 3, pp. 224- 274, 2001.

[14] K. LeFevre, D. DeWitt, and R. Ramakrishnan, "Mondrian Multidimensional K-Anonymity," Proc. 22nd Int'l Conf. Data Eng., pp. 25- 25, 2006.

[15] J. Friedman, J. Bentley, and R. Finkel, "An Algorithm for Finding Best Matches in Logarithmic Expected Time," ACM Trans. Mathematical Software, vol. 3, no. 3, pp. 209-226, 1977.

[16] A. Meyerson and R. Williams, "On The Complexity of Optimal k-Anonymity," Proc. 23rd ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems, pp. 223-228, 2004.

[17] G. Aggarwal, T. Feder, K. Kenthapadi, R. Motwani, R. Panigrahy, D. Thomas, and A. Zhu, "Approximation Algorithms for k-Anonymity," J. Privacy Technology, vol. 2005112001, pp. 1-18, 2005.

[18] R. Sandhu and Q. Munawer, "The Arbac99 Model for Administration of Roles," Proc. 15th Ann. Computer Security Applications Conf., pp. 229-238, 1999.

[19] E. Otoo, D. Rotem, and S. Seshadri, "Optimal Chunking of Large Multidimensional Arrays for Data Warehousing," Proc. ACM 10th Int'l Workshop on Data Warehousing and OLAP, pp. 25-32, 2007.

[20] W. Hoeffding, "On the Distribution of the Number of Successes in Independent Trials," The Annals of Math. Statistics, vol. 27, no. 3, pp. 713-721, 1956.

V.NEHA received B.Tech. in computer science and engineering from JNTUH Hyderabad and M.Tech. in computer science and Engineering from JNTUH Hyderabad, She is currently pursuing Mtech, Department of Computer Science and Engineering at St.peters Engineering college, Hyderabad, TS,INDIA.

KOLLURI DAVID RAJU is a Ph.D Research Scholar at the Rayalaseema University ,Kurnool, A.P, INDIA. He Received M.Tech degree in COMPUTER SCIENCE & ENGINEERING in the year 2010 and B.Tech degree in COMPUTER SCIENCE & ENGINEERING in the year 2002 from JNTUH Hyderabad, TS, INDIA and He Received Diploma in COMPUTER SCIENCE & ENGINEERING in the year 1996 from Kakatiya University ,Warangal, TS,INDIA. He is currently working As ASSOCIATE PROFESSOR, Department of COMPUTER SCIENCE & ENGINEERING at St.peters Engineering college ,Hyderabad, TS, INDIA. He is having 15 years of Teaching experience and 2 years of Industrial experience, published more than 10 papers in National/International Journals/Conferences. He is a Member of IEEE and IAENG(International association of Engineers), ISTE(Indian Society For Technical Education), . His areas of research include Data Mining & Data Warehousing, Big Data, Machine Learning ,Data Structures & Algorithms and Programming Languages.