ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

Further Investigations on Strategies Developed For Efficient Discovery of Matching Dependencies

R.Santhya[1], S.Latha[2], Prof.S.Balamurugan[3], S.Charanyaa[4]

Department of IT, Kalaignar Karunanidhi Institute of Technology, Coimbatore, TamilNadu, India[1,2,3]

Senior Software Engineer Mainframe Technologies Former, Larsen & Tubro (L&T) Infotech, Chennai, TamilNadu, India[4]

ABSTRACT: This paper details about various methods prevailing in literature for efficient discovery of matching dependencies. The concept of matching dependencies (MDs) has recently been proposed for specifying matching rules for object identification. Similar to the functional dependencies with conditions, MDs can also be applied to various data quality applications such as detecting the violations of integrity constraints. The problem of discovering similarity constraints for matching dependencies from a given database instance is taken into consideration. This survey would promote a lot of research in the area of information mining.

KEYWORDS: Data Anonymization, Matching Dependencies(MDs), Object, Similarity Constraints, Information Mining.

I. INTRODUCTION

Need for publishing sensitive data to public has grown extravagantly during recent years. Recent days have seen a steep rise in preserving data quality in the database community due to the huge amount of “dirty” data originated from different. These data often contain duplicates, inconsistencies and conflicts, due to various mistakes of men and machines. In addition to the cost of dealing with the huge volume of data, manually detecting and removing “dirty” data is definitely out of practice because human proposed cleaning methods may introduce inconsistencies again. Therefore, data dependencies, which have been widely used in the relational database design to set up the integrity constraints. Hence protecting privacy of individuals and ensuring utility of social network data as well becomes a challenging and interesting research topic.. In this paper we have made an investigation on the attacks by matching dependencies and possible solutions proposed in literature and efficiency of the same.

II. DISCOVERING MATCHING DEPENDENCIES

For identifying the abuse of integrity constraints and identification of duplicate objects, these data quality applications has been proposed as matching dependencies (MDs). The author studied the problem of identifying database instance. Firstly, they define the measures, support and confidence in order to evaluate the usage of MDs in the provided database instance. The discovery of MDs with particular usage requirements of support and confidence is also studied. Finally, the efficiency of the proposal methods is been evaluated experimentally.

III. DIFFERENTIAL DEPENDENCIES: REASONING AND DISCOVERY

In this paper, the author proposed a new algorithm known as differential dependencies(DDs) that specifies limitations on difference called differential functions. Among the various kinds of data’s like numerical values or text values, the importance of difference semantics is recently identified for declaring dependencies. A DDs shows that, if two tuples have distances on attributes X accepting with a particular differential functions on Y.

In this article, the author address various theoretical issues of differential dependencies. Then the author investigated the problem of how to discover DDs and differential keys from a provided dataset. The author demonstrated the discovery performance and the effectiveness of DDS in various real-world applications.

IV. ON DATA DEPENDENCIES IN DATA SPACES

In this paper the author defined common dependencies called a comparable dependency (CDs) that specifies the constraints on comparable attributes of data dependencies. This CDs overall covers the semantics of a huge class of dependencies in databases including FDs , metric FDs, MDs. Because, for the heterogeneous data in dataspaces called validation problem is to find whether a dependency holds in a data instance.

V. LEVERAGING MATCHING DEPENDENCIES FOR GUIDED USER FEEDBACK IN LINKED DATA APPLICATION

In this paper the author proposed a new approach for managing the integration quality and the feedback of the user for merging entity within application consuming linked open data.

Linked open data ease the publishing of large amount of structured data which enables the creation of a global data space on the web.

Based on the domain specific matching dependencies the term called utility measure is used to define the quality of a data space which contains multiple linked datasets.

In this paper the author presented a utility driven approach for identifying identity resolutions links from users. The contribution towards the paper is as follows,

1)Leveraging matching dependencies for consolidation of entities in application consuming linked data

2)A strategy for ranking identity resolution links based on approximated usage of expected user feedback.

3)Experimented evaluation of the proposed approach on real world and synthetic datasets.

The authors proposes (VPI-rules)ranks uncertain identity resolution links according to the potential benefits to the dataspace. The author also put forth his future work towards extending the proposed approach with other types of data quality constraints like comparable and order dependencies.

The author's said that there are many recent technologies and algorithm towards preservation of data mining. One of the methods is K-anonymity which should have at least k-1 tupple. Another major con about releasing the data is that the background knowledge will make the data unsafe. The author also designed the algorithm with low information loss. Gathering the information is a big task for the organizations to focus on such things. Different algorithms like decision tree, nearest neighbor method etc..are used for the limitation of knowledge discovery. In the recent years, the anonymization algorithm has been popularly known to all and the anonymization techniques are Generalization and Bucketization. The generalization is a principle which provides security and also increases the level of protection will inversely decrease the utilization level of this algorithm where the datum is stored in a common name. The attributes in the records can be splitted. Firstly, identifier which is used to uniquely identify the attributes like name, security number etc..Secondly, the sensitive attributes like phone number, zip code etc are been generalized. If this is been identified by the intruder then it will result in physical and emotional problems. Thirdly, Quasi Identifier (QI) is an identifier, when this data is been released, the possibility of getting the original data is high. Some of the encryption and decryption techniques with keys can be used to prevent the data. The disclosure of data to a certain level is not a problem but beyond the boundary level will leads to proximity breach.

The Cross – attribute correlation by FFD's can also bring vulnerability. No existing algorithms have prevented against FFD based privacy attacks. The author formalize the FFD privacy attack and also define the privacy model (d, l) – inference to combat the FD based attacks. Record linkage problem is also taken into account which can combine the quasi – identifiers with the information obtained from diverse external sources to re- identity the individual in released data. Against the record linkage problem an earliest principle k – anonymity in which each record is indistinguishable from at least k-1 from other records with respect to their QI and another improved principle. The algorithm 'l-diversity' is proposed but it also has some sensitive data. The Generalization is used to achieve both k-anonymity and l-diversity into groups. For the tuples in the same group their QI values are generalized to be identical so that they are indistinguishable to each other with regards to the QI values. These two principle k-anonymity and l-diversity have witnessed the proposals like t-closeness, (l, k) safety all these principles have also addressed this adversary knowledge by no one of them has considered the FFD.bThe author's have studied the problems of privacy publishing the microdata with FFD presenting with FFD as adversarial knowing. They defined as follows First, They defined FFD. Based on the impact of FFDs to privacy we distinguish "safe" FFDs that cannot able any FFD – based attack from the "unsafe" ones that can. Second, (d,l) – inference model to defend against the FFD - based attack. Thirdly, three novel grouping strategies to group sensitive values are explained. For each strategy, they analyze the amount of information loss by tupple suppression. Fourth, the study of the impact of multiple un-safe FFDs to anonymization is defined. We define efficient anonymization algorithms for multiple un-safe FFDs and measure both time performance and information loss of anonymization algorithm empirically. Lastly they have defined them by an expensive set of experiments; their results show that the anonymization of the microdata with low information loss is found when FFDs are available. The author also extended their work as follows by only considering numerical data, extended to cover categorical data and re-model their privacy frame work accordingly. Intersection grouping (IG) strategy will produce large amount of information loss which partition the sensitive values into groups that intersect in a chain. So, they proposed two strategy disjoint grouping (DG) which partitions the sensitive values into groups that not overlap and containment grouping (CG) which partition the sensitive values into groups that follow a strict containment relationship. Only they considered here about the single FFD.

They elaborate the reasoning to multiple FFDs and analyzed that some FFDs can be an representative of others. By this knowledge, they developed an anonymization algorithm for multiple unsafe FFDs and measure both the performance and information loss of the anonymization on algorithm empirically. The application scenario consists of two phase's data collection and publishing model. Firstly, in the data collection phase the data is collected by the publisher from the owner of the record. In the second phase which is publishing phase the collected data is provided to the third party service provider or the public by trusting them and with the knowledge of the data owners. In the top down approach, the set of all unique sensitive values are in unsafe FFD is found. Then we estimate the number of removed tuples, if it cannot be further divided into groups then it remains unchanged. Otherwise the values are splited into groups since because it reduces the information loss by suppression. In the bottom up approach, the sensitive attributes are splited into multiple groups and will combine the splited groups into disjoint set until the sets cannot be merged further with the help of IG, DG, and CG. Both the approaches should satisfy the (d, l) inference. In this paper the author studied about the privacy publishing of data that contains FFD and also formally defined the privacy model (d, l)-inference to prevent the disclosure that caused by FFDs .Finally they developed an adequate algorithm to anonymize the data with low information loss.

Protecting the macro data and micro data against unauthorized access has been a long goal of the database security. Several techniques and mechanisms are combined to serve this problem. The author's dealt with the problem of limiting disclosure sensitive rules by hiding some frequent item sets which frequently enough and finding them is the initial step towards association / correction rule or sequential patter mining. The high sensitive data's are the identified by the low sensitive data called "influence problem". Their problem is to reduce support of the rules in the given threshold and is referred as sanitization. In this paper the author explained and proved that identifying an optimal sanitization of the database is NP-hard. To overcome this problem Heuristic approach is used which sorts the item set based on their support and then tries to hide all of them in one-by-one fashion. In order to calculate the efficiency of the heuristic on another algorithm called cyclic algorithm is developed which also needs the item set to be hidden.

The author(s) has proposed a proposal which converts the original data sets into some unreal data sets such that any original data set is net able to reconstruct an unauthorized parts were steal some unauthorized parts were steal some portion of unrealized datasets, so there will be low probability of finding or relating with the original data. In this work the feature of new privacy preserving approach with 1D3 (Iterative dichotomister 3) which selects the test attribute based on the details obtained by the test outcome of decision tree algorithm. This ID3 will support only for discrete valued attributes. To support for continuous valued attributes C5.0 algorithm which splits the info based on the field which provides the maximum information gain.

VI. A GUIDED TOUR TO APPROXIMATE USING MATCHING

In this paper the author describes the problem of string matching issues. This is the very important issues in the growing area such as information retrieval and computational biology. In this paper number of experiments are presented for computing the different algorithms.

The string matching allowing errors also called as approximate string matching .The main goal of that is to performing the string matching in a text where both of them are corrupted.

The problem is finding the position of the text and allowing the limited numbers of errors in the matches .There are different errors are occur in the different applications.

In this survey the author presents the approximate string matching and focus on online searching. These will explaining the problem and its relevance, its statistical behaviour, history and current development and the central idea of the algorithm and complexity.

The problem depends upon the types of errors and the solution range from linear time to NP-Complete. In first references we focus on the problem appeared in a number of various field those time we mainly focus on the problem of computational biology, signal processing and text retrieval.

The 'N'number of applications for matching string is increases every day. So finding solution to the most problem based on the approximate matching string for instances hand writing recognition, Intrusion detection and virus .Image compression, data mining, optical character recognition ,file comparison and screen updating ,to name a few.

To understand all the application development necessarily need the important concept for that .So the basic knowledge on design and analysis of data structure and algorithm, basic text algorithm and formal languages. In this paper the dynamic programming algorithm is the first algorithm to solve this problem .It was rediscovered in the past , in different areas ,this algorithm computed the edit distance and it was latterly converted into a search algorithm .This is not well efficient , it is the flexible one to adopt the different distance function.

The next one is algorithm based on automate. This is also the old area this is the best worst case time algorithm and there is a time and space exponential on M and K.

The third one is Bit Parallelism algorithm based on parallelism of the computers when it works on bits.This is a new and active area. The main idea is parallelize the other algorithm using bits.So the outcome of this is practical point of view .It will works very efficiently .In this algorithm first find the element which is belonging to the other sections .The parallelism the work of the non deterministic automation that solves the problem and parallelize the work of the dynamic programming metric these two trends are used in that algorithm.

The fourth algorithm is filtering algorithm is simply fresh and very active. This algorithm filter the text, veryfastly discard the unmatch text area. In this paper the main goal is to present and explain the idea of the existing algorithm. The approximate string matching is a very active area.

VII. DIFFERENTIAL DEPENDENCIES REASONING AND DISCOVERY

In this paper the author describes the importance of the semantics like similarity and dissimilarity and declaring the dependencies among various types of data such as numerical or text values .The author propose novel differential dependencies which specifies the constraints on difference called differential functions.

In this paper first address the several issues of differential dependencies including subsumption order relation of differentialfunction, implication of dds, closure of a differential function, a sound and complete inference system and minimal cover for dds after that investigate a practical problem because of the hardness. We develop the discovery efficiency. Finally these demonstration to discover the effectiveness of dds in several application.

VIII. EFFICIENT CLUSTERING OF HIGH DIMENSIONAL DATA SETS WITH APPLICATION TO REFERENCE MATCHING

In this paper the author describes the problem involves clustering large datasets. It is computationally very expensive. The techniques for clustering when the database sets has either

- A limited number of clusters
- A low features dimensionality
- A small number of data points

In this paper the author presents the new techniques for clustering the large high dimensional data sets. The basic idea using a cheap, approximate distance measures are divide the data into subsets we call canopies. The canopies are used to solve the large clustering problem impossible become practical. using this canopies the output will be more accuracy. This will used in the many domains and uses the various techniques greedy agglomerative clustering, K-means etc.,.This will reduce the computational time.

Unsupervised clustering applied to many several important problems. The basic idea for performing the cluster in two ways. First divide the data into subsets called canopies, the expensive distance measurements are made among points that will occur in a normal canopy. Using cruder similarity measures to fastly form the canopies and then using the similarity measures to form smaller cluster So high speed and precision are obtained. In future work we are running this algorithm on the full set and expect to reduced the computation of five order of magnitude.

IX. CONCLUSION AND FUTURE WORK

This paper detailed about various methods prevailing in literature for efficient discovery of matching dependencies. The concept of matching dependencies (MDs) has recently been proposed for specifying matching rules for object identification. Similar to the functional dependencies (with conditions), MDs can also be applied to various data quality applications such as detecting the violations of integrity constraints. The problem of discovering similarity constraints for matching dependencies from a given database instance is taken into consideration. This survey would promote a lot of research in the area of information mining.

REFERENCES

1. Shaoxu Song, Lei Chen, "Efficient discovery of similarity constraints for matching dependencies", Data & Knowledge Engineering, Elsevier, 2013.
2. S. Abiteboul, R. Hull, V. Vianu, Foundations of Databases, Addison-Wesley, 1995.
3. R. Agrawal, T. Imielinski, A.N. Swami, Mining association rules between sets of items in large databases, SIGMOD Conference, 1993, pp. 207–216.
4. R. Bassée, J. Wijsen, Neighborhood dependencies for prediction, PAKDD, 2001, pp. 562–567.
5. C. Batini, M. Scannapieco, Data quality: concepts, methodologies and techniques, Data-Centric Systems and Applications, Springer, 2006.
6. L.E. Bertossi, S. Kolahi, L.V.S. Lakshmanan, Data cleaning and query answering with matching dependencies and matching functions, ICDT, 2011, pp. 268–279.
7. M. Bilenko, R.J. Mooney, W.W. Cohen, P. Ravikumar, S.E. Fienberg, Adaptive name matching in information integration, IEEE Intelligent Systems 18 (5) (2003) 16–23.
8. D. Bitton, J. Millman, S. Torgersen, A feasibility and performance study of dependency inference, ICDE, 1989, pp. 635–641.
9. P. Bohannon, W. Fan, F. Geerts, X. Jia, A. Kementsietsidis, Conditional functional dependencies for data cleaning, ICDE, 2007, pp. 746–755.
10. L. Bravo, W. Fan, F. Geerts, S. Ma, Increasing the expressivity of conditional functional dependencies without extra complexity, ICDE, 2008, pp. 516–525.
11. L. Bravo, W. Fan, S. Ma, Extending dependencies with conditions, VLDB, 2007, pp. 243–254.
12. T. Calders, R.T. Ng, J. Wijsen, Searching for dependencies at multiple abstraction levels, ACM Transactions on Database Systems 27 (3) (2002) 229–260.
13. F. Chiang, R.J. Miller, Discovering data quality rules, PVLDB 1 (1) (2008) 1166–1177.
14. W.W. Cohen, Integration of heterogeneous databases without common domains using queries based on textual similarity, SIGMOD Conference, 1998, pp. 201–212.
15. G. Cong, W. Fan, F. Geerts, X. Jia, S. Ma, Improving data quality: consistency and accuracy, VLDB, 2007, pp. 315–326.
16. A.K. Elmagarmid, P.G. Ipeirotis, V.S. Verykios, Duplicate record detection: a survey, IEEE Transactions on Knowledge and Data Engineering 19 (1) (2007) 1–16.
17. W. Fan, Dependencies revisited for improving data quality, PODS, 2008, pp. 159–170.
18. W. Fan, H. Gao, X. Jia, J. Li, S. Ma, Dynamic constraints for record matching, The VLDB Journal (2010) 1–26.

19. W. Fan, F. Geerts, L.V.S. Lakshmanan, M. Xiong, Discovering conditional functional dependencies, ICDE, 2009, pp. 1231–1234.
20. W. Fan, J. Li, X. Jia, S. Ma, Reasoning about record matching rules, PVLDB, 2009.
21. W. Fan, J. Li, S. Ma, N. Tang, W. Yu, Interaction between record matching and data repairing, SIGMOD Conference, 2011, pp. 469–480.
22. W. Fan, S. Ma, Y. Hu, J. Liu, Y. Wu, Propagating functional dependencies with conditions, PVLDB 1 (1) (2008) 391–407.
23. P.A. Flach, I. Savnik, Database dependency discovery: a machine learning approach, AI Communications 12 (3) (1999) 139–160.
24. J. Gardezi, L.E. Bertossi, I. Kiringa, Matching dependencies with arbitrary attribute values: semantics, query answering and integrity constraints, LID, 2011, pp. 23–30.
25. C. Giannella, E.L. Robertson, On approximation measures for functional dependencies, Information Systems 29 (6) (2004) 483–507.
26. L. Golab, H.J. Karloff, F. Korn, A. Saha, D. Srivastava, Sequential dependencies, PVLDB 2 (1) (2009) 574–585.
27. L. Golab, H.J. Karloff, F. Korn, D. Srivastava, B. Yu, On generating near-optimal tableaux for conditional functional dependencies, PVLDB 1 (1) (2008) 376–390.
28. L. Gravano, P.G. Ipeirotis, N. Koudas, D. Srivastava, Text joins in an rdbms for web data integration, WWW, 2003, pp. 90–101.
29. Y. Huhtala, J. Kärkkäinen, P. Porkka, H. Toivonen, Efficient discovery of functional and approximate dependencies using partitions, ICDE, 1998, pp. 392–401.
30. Y. Huhtala, J. Kärkkäinen, P. Porkka, H. Toivonen, Tane: an efficient algorithm for discovering functional and approximate dependencies, The Computer
31. I.F. Ilyas, V. Markl, P.J. Haas, P. Brown, A. Aboulnaga, Cords: automatic discovery of correlations and soft functional dependencies, SIGMOD Conference, 2004, pp. 647–658.
32. R.S. King, J.J. Legendre, Discovery of functional and approximate functional dependencies in relational databases, JAMDS 7 (1) (2003) 49–59.
33. J. Kivinen, H. Mannila, Approximate inference of functional dependencies from relations, Theoretical Computer Science 149 (1) (1995) 129–149.
34. N. Koudas, A. Saha, D. Srivastava, S. Venkatasubramanian, Metric functional dependencies, ICDE, 2009, pp. 1275–1278.
35. S. Kramer, B. Pfahringer, Efficient search for strong partial determinations, KDD, 1996, pp. 371–374.
36. S.E. Madnick, H. Zhu, Improving data quality through effective use of data semantics, Data & Knowledge Engineering 59 (2) (2006) 460–475.
37. H. Mannila, K.-J. Räihä, Design of Relational Databases, Addison-Wesley, 1992.
38. H. Mannila, K.-J. Räihä, Algorithms for inferring functional dependencies from relations, Data & Knowledge Engineering 12 (1) (1994) 83–99.
39. A. McCallum, K. Nigam, L.H. Ungar, Efficient clustering of high-dimensional data sets with application to reference matching, KDD, 2000, pp. 169–178.
40. G. Navarro, A guided tour to approximate string matching, ACM Computing Surveys 33 (1) (2001) 31–88.
41. B. Pfahringer, S. Kramer, Compression-based evaluation of partial determinations, KDD, 1995, pp. 234–239.
42. T. Scheffer, Finding association rules that trade support optimally against confidence, Intelligent Data Analysis 9 (4) (2005) 381–395.
43. J.C. Schlimmer, Efficiently inducing determinations: a complete and systematic search algorithm that uses optimal pruning, ICML, 1993, pp. 284–290.
44. S. Song, L. Chen, Discovering matching dependencies, CIKM, 2009, pp. 1421–1424.
45. S. Song, L. Chen, Differential dependencies: reasoning and discovery, ACM Transactions on Database Systems 36 (4) (2011).
46. S. Song, L. Chen, P.S. Yu, On data dependencies in dataspaces, ICDE, 2011, pp. 470–481.
47. U. ul Hassan, S. O'Riain, E. Curry, Leveraging matching dependencies for guided user feedback in linked data applications, Proceedings of the Ninth International Workshop on Information Integration on the Web, IIWeb '12, ACM, New York, NY, USA, 2012, pp. 5:1–5:6.
48. H. Wang, R. Liu, Privacy-preserving publishing microdata with full functional dependencies, Data & Knowledge Engineering 70 (3) (2011) 249–268.
49. C.M. Wyss, C. Giannella, E.L. Robertson, Fastfds: a heuristic-driven, depth-first algorithm for mining functional dependencies from relation instances —extended abstract, DaWaK, 2001, pp. 101–110.
50. B.Powmeya , Nikita Mary Ablett ,V.Mohanapriya,S.Balamurugan,"An Object Oriented approach to Model the secure Health care Database systems,"In proceedings of International conference on computer , communication & signal processing(IC^{3}SP)in association with IETE students forum and the society of digital information and wireless communication,SDIWC,2011,pp.2-3
51. Balamurugan Shanmugam, Visalakshi Palaniswami, "Modified Partitioning Algorithm for Privacy Preservation in Microdata Publishing with Full Functional Dependencies", Australian Journal of Basic and Applied Sciences, 7(8): pp.316-323, July 2013
52. Balamurugan Shanmugam, Visalakshi Palaniswami, R.Santhya, R.S.Venkatesh "Strategies for Privacy Preserving Publishing of Functionally Dependent Sensitive Data: A State-of-the-Art-Survey", Australian Journal of Basic and Applied Sciences, 8(15) September 2014.
53. S.Balamurugan, P.Visalakshi, V.M.Prabhakaran, S.Chranyaa, S.Sankaranarayanan, "Strategies for Solving the NP-Hard Workflow Scheduling Problems in Cloud Computing Environments", Australian Journal of Basic and Applied Sciences, 8(15) October 2014.
54. Charanyaa, S., et. al., , A Survey on Attack Prevention and Handling Strategies in Graph Based Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 2(10): 5722-5728, 2013.
55. Charanyaa, S., et. al., Certain Investigations on Approaches forProtecting Graph Privacy in Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 1(8): 5722-5728, 2013.
56. Charanyaa, S., et. al., Proposing a Novel Synergized K-Degree L-Diversity T-Closeness Model for Graph Based Data Anonymization. International Journal of Innovative Research in Computer and Communication Engineering, 2(3): 3554-3561, 2014.
57. Charanyaa, S., et. al., , Strategies for Knowledge Based Attack Detection in Graphical Data Anonymization. International Journal of Advanced Research in Computer and Communication Engineering, 3(2): 5722-5728, 2014.
58. Charanyaa, S., et. al., Term Frequency Based Sequence Generation Algorithm for Graph Based Data Anonymization International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.

59. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Certain Investigations on Strategies for Protecting Medical Data in Cloud", International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014
60. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Investigations on Remote Virtual Machine to Secure Lifetime PHR in Cloud ", International Journal of Innovative Research in Computer and Communication Engineering Vol 2, Issue 10, October 2014
61. V.M.Prabhakaran, Prof.S.Balamurugan, S.Charanyaa," Privacy Preserving Personal Health Care Data in Cloud" , International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014
62. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, "Investigations on Evolution of Strategies to Preserve Privacy of Moving Data Objects" International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.
63. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Certain Investigations on Securing Moving Data Objects" International Journal of Innovative Research in Computer and Communication Engineering, 2(2): 3033-3040, 2014.
64. P.Andrew, J.Anish Kumar, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Survey on Approaches Developed for Preserving Privacy of Data Objects" International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014
65. S.Jeevitha, R.Santhya, Prof.S.Balamurugan, S.Charanyaa, " Privacy Preserving Personal Health Care Data in Cloud" International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 2, October 2014.
66. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "Investigations on Methods Evolved for Protecting Sensitive Data", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, Decermber 2014.
67. K.Deepika, P.Andrew, R.Santhya, S.Balamurugan, S.Charanyaa, "A Survey on Approaches Developed for Data Anonymization", International Advanced Research Journal in Science, Engineering and Technology Vol 1, Issue 4, Decermber 2014.
68. S.Balamurugan, S.Charanyaa, "Principles of Social Network Data Security" LAP Verlag, Germany, ISBN: 978-3-659-61207-7, 2014
69. S.Balamurugan, S.Charanyaa, "Principles of Scheduling in Cloud Computing" Scholars' Press, Germany,, ISBN: 978-3-639-66950-3, 2014
70. S.Balamurugan, S.Charanyaa, "Principles of Database Security" Scholars' Press, Germany, ISBN: 978-3-639-76030-9, 2014

BIOGRAPHY

R.Santhya and S.Latha are currently pursuing their B.Tech. degree in Information Technology at KalaignarKarunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India. Their areas of research interests include Network Security, Cloud Computing and Database Security.

Prof.S.Balamurugan obtained his B.Tech degree in Information Technology from P.S.G. College of Technology, Coimbatore, Tamil Nadu, India and M.Tech degree in Information Technology from Anna University, Tamil Nadu, India respectively. He is currently working towards his PhD degree in Information Technology at P.S.G. College of Technology, Tamil Nadu, India. At present he holds to his credit **65 papers International Journals and IEEE/ Elsevier International Conferences.** He is currently working as Assistant Professor in the Department of Information Technology, Kalaignar Karunanidhi Institute of Technology, Coimbatore, Tamil Nadu, India affiliated to Anna University TamilNadu, India. He is **State Rank holder** in schooling. He was **University First Rank holder** M.Tech. Semester Examinations at Anna University, Tamilnadu, India. He served as a Joint Secretary of IT Association, Department of Information Technology, PSG College of Technology, Coimbatore, Tamilnadu, India. He is the **recipient of gold medal and certificate of merit** for best journal publication by his host institution **consecutively for 3 years**. Some of his professional activities include invited Session Chair Person for two Conferences. He has guided 16 B.Tech projects and 2 M.Tech. projects. He has won a best paper award in International Conference. His areas of research interest accumulate in the areas of Data Privacy, Database Security, Object Modeling Techniques, and Cloud Computing. He is a life member of ISTE,CSI. **He has authored a chapter in an International Book "Information Processing" published by I.K. International Publishing House Pvt. Ltd, New Delhi, India, 978-81-906942-4-7. He is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**

S.Charanyaa obtained her **B.Tech** degree in Information Technology and her **M.Tech** degree in Information Technology from Anna University Chennai, Tamil Nadu, India. She was **gold medalist** in her B.Tech. degree program. She has to her credit **27 publications in various International Journals and Conferences**. Some of her outstanding achievements at school level include **School First Rank holder** in **10th and 12th grade**. She was working as Software Engineer at Larsen & Turbo Infotech, Chennai for 3 years where she got promoted as Senior Software Engineer and worked for another 2 years. She worked at different verticals and worked at many places including Denmark, Amsderdam handling versatile clients. She is also the recipient of **best team player award for the year 2012 by L&T**. Her areas of research interest accumulate in the areas of Database Security, Privacy Preserving Database, Object Modeling Techniques, and Cloud Computing. **She is the author of 3 books titled "Principles of Social Network Data Security", ISBN: 978-3-659-61207-7, "Principles of Scheduling in Cloud Computing" ISBN: 978-3-639-66950-3, and "Principles of Database Security", ISBN: 978-3-639-76030-9.**