International Conference on Recent Innovations in Engineering and Management
Dhananjay Mahadik Group of Institutions (BIMAT) Kolhapur, Maharashtra (ICRIEM-16)
23rd March 2016, www.conferenceworld.in ISBN: 978-81-932074-5-1

STATE OF AFFAIRS DEPENDENT SAFEKEEPING DEPLOYMENT TOWARDS THE OUTSOURCED SYSTEM INFRINGEMENT AND TO UNEARTH THE ODDS OF THE INTERLOPER TO COMPROMISE THE SHELTERED SPECIFICS AS WELL AS BREACH DISCRETION

Ms. Jyoti Shirahatti[1], Prof. Ramesh Ranadive[2], Dr. Suryakant B Patil[3]

[1,2]Dr. D. Y. Patil Polytechnic, Bawada, Kolhapur, MH, (India)
[3] Sanjay Ghodawat Gr. of Institutions, Atigre, Kolhapur, MH, (India)

ABSTRACT

In this paper we have proposed an idea for state of affairs dependent safekeeping deployment towards the outsourced system infringement and to unearth the odds of the interloper to compromise the sheltered specifics as well as breach discretion.The Data Distributors organizations higher the third party to outsource their sensitive data.The aim is to unearth when their sensitive data has been leaked by some agents, and if possible to identify those guilty agents. A case study scenario where the original sensitive data cannot be perturbed is studied; in some cases it is imperative not to alter or modify the original distributor's data. In such circumstances, one cannot modify the data, so one has to go by the methodology in which data alteration is not required. In this paper we have proposed a solution under such versatile different circumstances. The objective is to find out the probability that the leaked data came from the agents as opposed to other reliable sources. Not only to this system wants to estimate the probability the agents leaked data, but it would also like to find out if one of them in a specific was more prone towards the data leakage. The data allocation methods enable the distributor "intelligently" give data to agents. Fake objects can be added to unearth the guilty agent, to address this problem different circumstances are studied. Based on which the data request triggered; the classification of data request either Explicit Data Request or Sample Data Request and sub categories would be with or without the fake objects are inserted.

Keywords: Infringement, Unearth Odds, Interloper, Breach Discretion, Data Leakage, Fake Object, Methodology, Confidentiality, Guilty Agents

I. INTRODUCTION

Guilty Agents are the agents who leak the sensitive information and specifics of the distributor to the unauthorized third party. Perturbation is a method in which the original data is altered and made somewhat less sensitive so that it will provoke itself as less useful data to some intruder. Fingerprinting [18] and Watermarking [19] are such perturbation techniques. Instead of these techniques, easy methods like changing exact data values by ranges can be

done under perturbation. However, in some cases it is important not to modify the original distributor's data. For example, the data of any scientific research cannot be altered by even a single margin. In such circumstances, one cannot use perturbation techniques, so he or she has to go by the method in which data alteration is not required. This proposed system provides a solution under such circumstances. For example, a hospital may give some patients information to some researchers for developing a new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various othersoutsource companies. The trusted third parties are known as the agents. For example, if an outsourcer is doing the payroll, he must have the exact salary and customer bank account numbers, or if medical researchers will be treating patients (as opposed to simply computing statistics), they may need accurate values for the patients.For example, a company may have partnerships with others that require sharing customer data. Another enterprise may outsource its data processing, so information must be given to various other companies. The trusted third parties are called the agents. The goal is to unearth when the distributor's sensitive data has been compromised by agents, and if possible to identify the agent responsible for the same [17].

II. LITERATURE SURVEY

Sethuram has mentioned in his paper "Data Loss/Leakage Prevention" that in today's business world, many organizations use Information Systems to manage their sensitive and business critical information. The need to protect such a key component of the organization cannot be over emphasized. Leakage unearthion system has proposed by Panagiotis Papadimitriou; Hector Garcia-Molina [1] which can enable us to unearth the guilty leaker without changing the integrity of the original data. P.Barge has mentioned in his paper "A Novel Data Leakage Detection" that his objective is to be able to unearth an agent who leaks any portion of his data by simple manipulation of the data by shifting the LSB.Papadimitriou has mentioned in paper "Data Leakage Detection" that some of the data distributed may be found in unauthorized place so distributor must asses and find guilty agents and data leakage.Amol Gharpande has mentioned in paper "Data leakage detection" that In both the commercial and defense sectors a compelling need is emerging for rapid, yet secure, dissemination of Information.Rekha Jadhav has mentioned in her paper "Data Leakage detection" that Perturbation is a very useful technique where the data is altered and made less sensitive 'before being handed to agents.Vaidhya Archana has mentioned in her paper "Data Leakage detection" that Modern business activities rely on extensive email exchange. Email leakages have become widespread, and the severe damage caused by such leakages constitutes a disturbing problem for organizations.AishwaryaPotdar has mentioned in her paper "Data Leakage Detection in networks". In the field of business, the owner of any organization, company or business firm having some crucial data may need to share it with third-parties. These trusted third-parties may use this data for their own benefit causing reputational and monetary damage to the owner's company.Dube Krishna has mentioned in his paper "Detection of data leakage using Fake data" that As organizations increase their reliance on, presumably distributed, data systems for daily

business, they become additional vulnerable to security breaches whilst they gain productivity and potency blessings.

They have used some allocation strategies to help find out guilty person. Chuanxian Jiang; Xiaowei Chen; ZhiLi[2] has given the feasibility of embedded watermarks on relational databases in Discrete Wavelet transform domain. They have proposed that by By employing linear correlation detecting method, a watermark can be embedded into relational database in DWT domain. Marecki; MudhakarSrivatsa; PradeepVarakantham[3] has formulated data leakage prevention problem as a Partially Observable Markov Decision Processes. They show how to embed digital watermarking in to the information and derive optimal information sharing strategies for the sender and optimal information leakage strategies for rational-malicious recipient. Yingjiu Li, Member; VipinSwarup; SushilJajodia[4] has presented a for fingerprinting relational data by extending watermarking scheme. The primary new capability provided by this technique is that, under reasonable assumptions, it can embed and unearth arbitrary bit string marks in relation. Mohamed Shehab; Elisa Bertino; ArifGhafoor[5] has formulated the watermarking of relational databases as a constrained optimization problem and has discussed efficient techniques to handle the constraint. Preeti Patil, NitinChavan, Srikantha Rao and S B Patil[10] present the probability evaluations during Building of a Secure Data Warehouse by Enhancing the ETL Processes for Data Leakage.Preeti Patil, Srikantha Rao, S B Patil[8] discussed the heterogeneity Problem in the ETL followed by the improvement in the security. Preeti S Patil, Srikantha Rao and Suryakant B Patil[9]presented the Optimization of Data Warehousing System by Simplification in Reporting and Analysis. P. S. Patil, S. Rao and S. B. Patil[11] discussed the Data integration problem of structural and semantic heterogeneity with data warehousing framework models.

III. PROPOSED METHODOLOGY - EXPERIMENTATION AND RESULTS:

Diversified Circumstances were considered for enhancement of Security by unearthing the Guilty Agent in addition to the Leakage of the Data during legal Data Transfer. Considering two scenarios Explicit Requests as well as Sample Data Requests; the Experimentations and Results are performed on the set of data.

3.1 Experimentation and Results for Sample Data Requests

In Sample Data Request approach as the agents are not interested inparticular objects, the object sharing is not required to be explicitlydefined by respective requests. Distributor is insisted toassign specific objects to multiple agents only if thenumber of requested objectsexceeds the numberof objects in set. The more hard it is to unearth a guilty agent if the more data objects the agentsrequest in total, the more recipients on an average an objecthas; and the more objects are shared among allagents.

File No.	without FO	with FO
1	25.3	39.4
2	23.4	33.4
3	27.4	42.5
4	24.9	34.2
5	26.4	27.2
6	28.4	37.6

Table 1: Results for Sample Data Requests

In two different ways thisnumberchanges: (a) considering that the number of agents is fixedand vary their sample sizes, (b) vary the number ofagents who send request for data. The later one captures howa real issue may get develop. The distributor may act toattract less or more agents for his or her data, but he or she does nothave any control upon agents' requests. We canalso increase randomlythe value of the load, as enhancing in the number of agents allows while varying agents' requestsposes an upper bound [5].

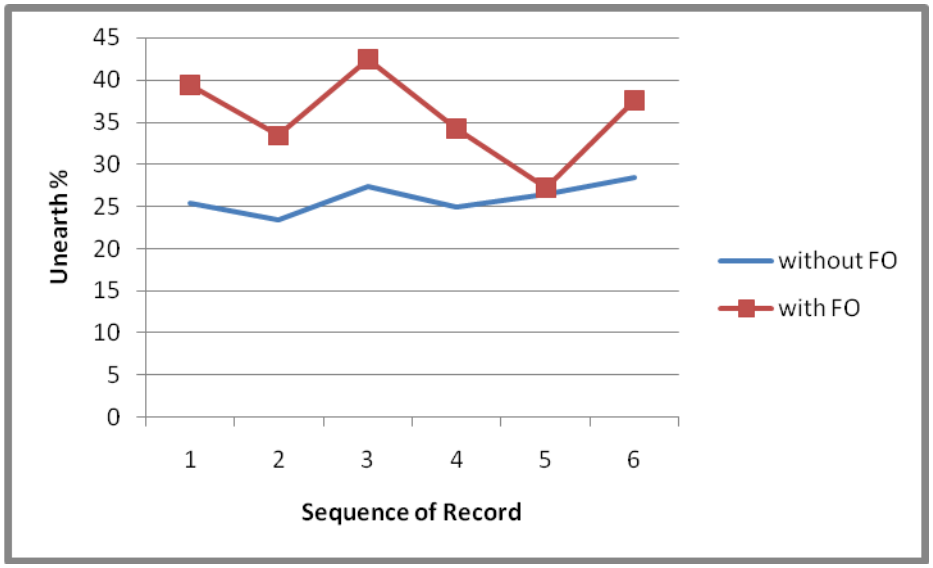

Figure 1:Sample Data Requests

3.2 Experimentaion and Results for Explicit Requests

Explicit Data Requests Experimentation in the first place, the goal was tosee whether fake objects in the distributed data sets yieldmajor enhancement in our chances of detecting aguilty agent.

File No.	without FO	with FO
1	35.2	46.8
2	26.4	48.2
3	28.6	45.2
4	35.6	42.8
5	32.6	46.5
6	35.6	48.4

Table 2: Results for Explicit Data Requests

In the second place, we wanted to assess the explicit data requests without fake objects for the same. Our center of attention was on scenarios with a small number of objects that areshared among multiple agents. These are the most excitingscenarios, since object sharing makes it hardto differentiate a guilty from non-guilty agents. Scenarioswith additional objects to distribute or scenarios with objectsshared among fewer agents are apparently easier to handle [1].

The Scenarios studied in the earlier work were Building of a Secure Data Warehouse [1], Data Extraction, Transformation and Loading [5], Integrated ERP System For Improving the Functional Efficiency of the Organization [7], Improvement in the efficiency of web based search engines [8], Network Traffic Optimization for Performance Improvement in the Web Service Infrastructures [9],Network architecture and design for optimized web page clustering with customized local proxy server [10], Copy right of the revised architecture for improvement in the design [11], Architecture for vigorous GSM handovers in Mobile cellular networks [12],

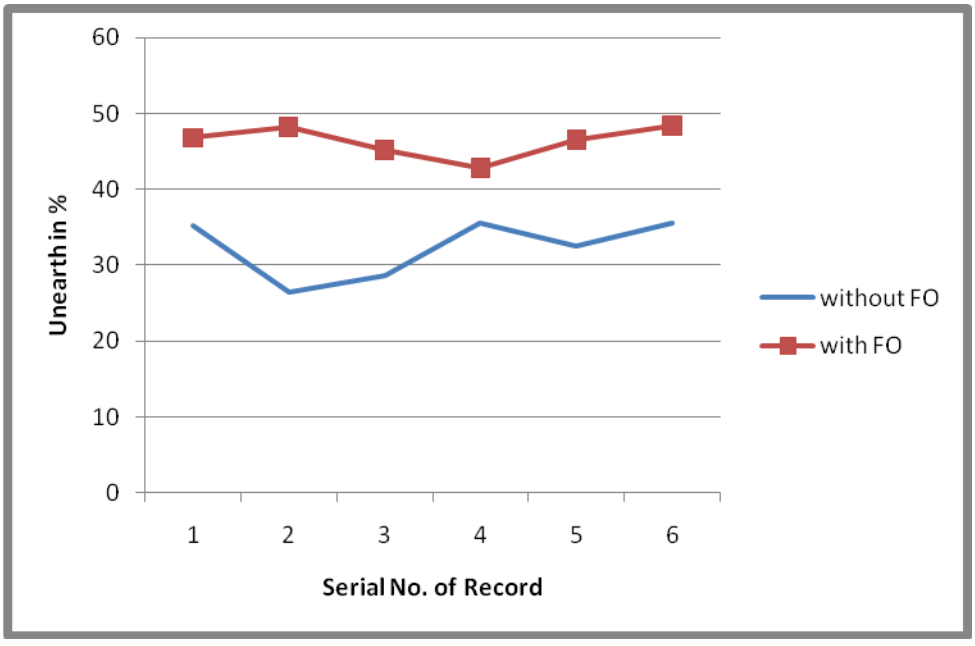

Figure 2: Explicit Data Requests

Optimization of Data Warehousing System[16],Enhanced Software Development Strategy Implying High Quality Design for Large Scale Database [20], Data integration problem of structural and semantic heterogeneity [23].As far as scenarios with several objects to distributeand many overlapping agent requests are apprehensive,they are alike to the scenarios we study, since we canmap them to the distribution of several tiny subsets.

IV. CONCLUSION

In spite of several hurdles, with experimentations we have shown it is possibleto assess the likelihood that an agent is responsible in the state of affairs dependent safekeeping deployment towards the outsourced system infringement and to unearth the odds of the interloper to compromise the sheltered specifics as well as breach discretion successfully. Due to fake object methodology guilty agents can be found very easily and fake object methodology is easier to implement as compared to any other methodology. Due to this methodology we can find who has leaked the sensitive data and from where it is leaked. In the experimentations data is given with addition of fake object to more than one person and data is assessed by the person who have distributed data to agents and on basis of the fake object provided to all agents data leakage and guilty agent is unearthed.In a perfect world there would be no need to handover sensitive data to agents that may unknowingly ormaliciously leak it. And even if we had to hand over sensitivedata, in a perfect world we could watermark eachobject so that we could trace its origins with absolutecertainty. However, in many cases we must indeed workwith agents that may not be 100% trusted, and we maynot be certain if a leaked object came from an agent orfrom some other source, since certain data cannot admitwatermarks.Our future work includes an investigation of models for cases where agents can collude and unearth fake tuples? Further one more issue is an extension of our allocation methods so that they cantake care agent requests in an online fashion; the presentedstrategies consider that there is a fixed set of agents with requests known well in advance

REFERENCES

[1] Preeti Patil, NitinChavan, Srikantha Rao and S B Patil; “Building of a Secure Data Warehouse by Enhancing the ETL Processes for Data Leakage”. International Journal of Computer Applications IJCA2012 Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET 2012): Volume- II, March , 2012 ISBN: 973-93-80864-48-5 Foundation of Computer Science, New York, USA 18-23

[2] Papadimitriou, Panagiotis, and Hector Garcia-Molina. "Data leakage detection." *Knowledge and Data Engineering, IEEE Transactions on* 23.1 (2011): 51-63.

[3] Gharpande, Amol O., and Ms VM Deshmukh. "Data Leakage Detection." *International Journal Of Computer Science And Applications* 6.2 (2013).

[4] Jadhav, Rekha. "Data Leakage Detection." *International Journal of Computer Science & Communication Networks* 3.1 (2013): 37.

[5] Preeti Patil, Srikantha Rao, S B Patil; "Data Extraction, Transformation and Loading by semantic web technologies and ontology for heterogeneity Problem" International Journal of Computer Science and Applications (IJCSA 2011 Issue-II), ISSN 0974-0767.

[6] Barge, Priyanka, PratibhaDhawale, and NamrataKolashetti. "A Novel Data Leakage Detection." *International Journal of Modern Engineering Research* 3.1 (2013).

[7] Vaidya, Archana, et al. "Data Leakage Detection." *International Journal of Advances in Engineering & Technology, ISSN* (2012)

[8] S B Patil, Preeti Patil, Vijay Surywanshi; "Integrated ERP System For Improving the Functional Efficiency of the Organization by Customized Architecture", International Journal of Computer Engineering & Applications, ISSN 23221-3469, Volume IV, Issue III, May 2014, Pages112-120. (Journal Impact Factor: 2.849)

[9] S B Patil, Preeti Patil, "Improvement in the efficiency of web based search engines by increasing the page rank based on referring factors". In international Journal of Information Technology & Management Information System (IJITMIS), volume 5, Issue 1, January - April (2014), pp. 53-59. Journal Impact Factor (2014): 6.2217 Calculated by GISI

[10] S B Patil, Preeti Patil, "Network Traffic Optimization for Performance Improvement inthe Web Service Infrastructures By Categorization Of The Web Contents With Size Reduction Approach". In International Journal of Advanced Research in Engineering & Technology (IJARET), Volume 5, Issue 4, April (2014), pp. 198-204 Journal Impact Factor (2014): 7.8273 Calculated by GISI

[11] S B Patil, Preeti Patil, "Network architecture and design for optimized web page clustering with customized local proxy server to reduce user-perceived latency and network resource requirements in the world wide web". In International Journal of Computer Engineering & Technology (IJCET), Volume 5, Issue 4, April (2014), pp. 210-217. Journal Impact Factor (2014): 8.5328

[12] S B Patil, Preeti Patil, "Copy right of the revised architecture for improvement in the design of data structure of the tree to enhance the applicability" in International Journal of Intellectual Property Rights (IJIPR), Volume 5, Issue 1, Jan - June (2014), pp. 22-29.

[13] S B Patil, Preeti Patil, "Architecture for vigorous GSM handovers in Mobile cellular networks by Jettisoning Hidden Terminal Problem", International Journal of Computer Engineering & Applications, ISSN 23221-3469, Volume VI, Issue III, June 2014, Pages121-131. (Journal Impact Factor: 2.849)

[14] Potdar, MsAishwarya, et al. "Data Leakage Detection In Networks."

[15] Dube, Krishna, B. Sailaja, and Ravi Mathey. "Detection of Data Leakage Using Fake Data."

[16] Sethuraman, Hariharan, and Mohammed Abdul Haseeb. "Data Loss/Leakage Prevention." (2012).

[17] Preeti S Patil, Srikantha Rao and Suryakant B Patil; "Optimization of Data Warehousing System: Simplification in Reporting and Analysis". International Journal of Computer Applications IJCA 2011 Proceedings on International Conference and workshop on Emerging Trends in Technology, 2011. Published by Foundation of Computer Science ISSN 0975 – 8887, (9):33-37, 2011.

[18] JanuszMarecki, MudhakarSrivatsa, PradeepVarakantham. 2007. A Decision Theoretic Approach to Data Leakage Prevention. IEEE International Conference on Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust.

[19] Yingjiu Li, VipinSwarup, SushilJajodia. 2005. Fingerprinting Relational Databases: Schemes and Specialties. IEEE Transaction

[20] Mohamed Shehab; Elisa Bertino, ArifGhafoor. 2008. Watermarking Relational Databases Using Optimization-Based Techniques. IEEE Transactions on Knowledge and Data Engineering.

[21] S. B. Patil, SamadhanSonavane, Srikantha Rao, Preeti Patil; "Enhanced Software Development Strategy Implying High Quality Design for Large Scale Database Projects", International Conference & Workshop on Emerging Trends in Technology ICWET'12, ISBN 978-0-615-58717-2, February 24–25, 2012, Pages: 478-483

[22] Yuer Wang, Zhongjie Zhu, Feng Liang, Gangyi Jiang. 2008. Watermarking Relational Data Based on Adaptive Mechanism. Proceedings of IEEE International Conference on Information and Automation.

[23] Simitsis, A.; Vassiliadis, P.; Sellis, T. "Optimizing ETL processes in data warehouses", Data Engineering 21st International Conference, 2005. ICDE 2005. Proceedings. Publication Year: 2005, Page(s): 564 – 575.

[24] P. S. Patil, S. Rao and S. B. Patil; "Data integration problem of structural and semantic heterogeneity: data warehousing framework models for the optimization of the ETL processes", ICWET '11 Proceedings of the International Conference & Workshop on Emerging Trends in Technology, ACM New York, NY, USA ©2011, ISBN: 978-1-4503-0449-8, Page(s): 500-504