Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.593 – 596

RESEARCH ARTICLE

TRACKING DOWN DATA LEAKAGE BY USING ENCRYPTED FAKE OBJECTS

Anugu Rahul Reddy[1], BalaKrishna Bangaru[2]

[1]M.Tech 2nd Year, Dept. of CSE, JNTU, HYDERABAD, INDIA

[2]Assistant Professor, Dept. of CSE, JNTU, HYDERABAD, INDIA

[1] anugu.rahulreddy@gmail.com; [2] balakrishna.bangaru@gmail.com

Abstract - This paper contains concept of data leakage, its causes of leakage and different techniques to protect and detect the data leakage. The value of the data is incredible, so it should not be leaked or altered. In the field of IT, huge database is being used. This database is shared with multiple people at a time. But during this sharing of the data, there are huge chances of data vulnerability, leakage or alteration. So, to prevent these problems, a data leakage detection system has been proposed. This paper includes brief idea about data leakage detection and a methodology to detect the data leakage persons.

I. INTRODUCTION

In the case of doing business, sometimes sensitive data must be handed over to trusted third parties. The distributor can assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. Using an analogy with cookies stolen from a cookie jar, if we catch Freddie with a single cookie, he can argue that a friend gave him the cookie. If the distributor sees ‘enough evidence´ that an agent leaked data, he may stop doing business with him, or may initiate legal proceedings. In this paper we develop a model for assessing the ‘guilt´ of agents. We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker. Finally, we also consider the option of adding ‘fake´ objects to the distributed set. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of watermark for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty.

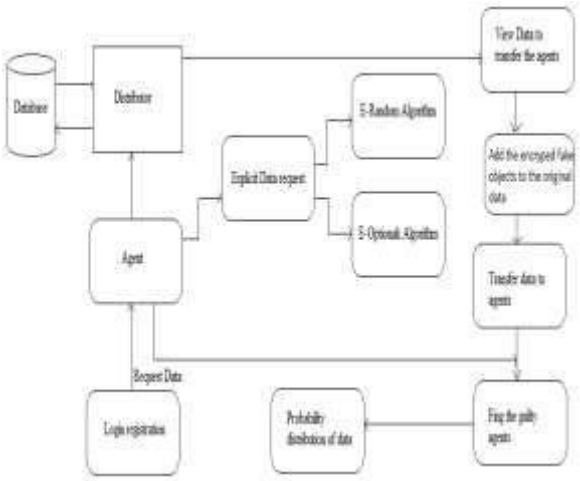

FIGURE 1: System Architecture Block Diagram

II. EXISTING SYSTEM

In existing system data leakage detection is handled by watermarking, e.g., a unique code is embedded in each distributed copy. If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified.

Watermarks can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious. E.g. A hospital may give patient records to researchers who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Many times agent get to know that the data will be watermark that time the data will be erase by the agent that time distributor never knows that who is the leaker.

Another enterprise may out source its data processing, so data must be given to various other companies. We call the owner of the data the distributor and the supposedly trusted third parties the agents. In many cases distributor must indeed work with agents that may not be trusted, and distributor may not be sure that a leaked object came from an agent or from some other source, since sure data cannot admit watermarks. In existing system there is few problem like fixed agents and existing system work comparable with agents whose request known in advance. Also with adding fake object original sensitive data cannot be alter and absences of agent guilt models that capture leakage scenarios and appropriate model for cases where agents can collude and identify fake tuples. Lastly system is not online capture of leak scenario also in existing system more focus on data allocation problem.

DISADVANTAGE:

This data is vulnerable to attacks. There are several techniques by which the watermark can be removed. Thus the data will be vulnerable to attacks.

III. PROPOSED SYSTEM

In this we propose to develop a model for assessing the guilt of agents. We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker. Finally, we also consider the option of adding fake objects to the distributed set. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of watermark for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty.

ADVANTAGE:

This system includes the data hiding along with the provisional software with which only the data can be accessed. This system gives privileged access to the administrator (data distributor) as well as the agents registered by the distributors. Only registered agents can access the system. The user accounts can be activated as well as cancelled. The exported file will be accessed only by the system. The agent has given only the permission to access the software and view the data. The data can be copied by our software. If the data is copied to the agent' system the path and agent information will be sent to the distributors email id thereby the identity of the leaked user can be traced

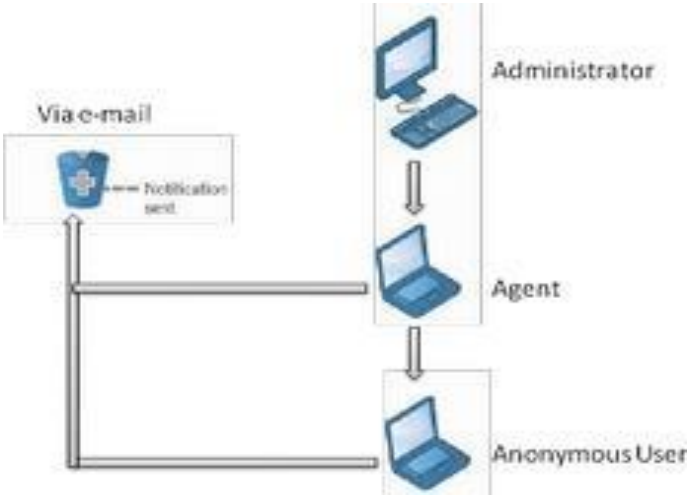

Figure 1 **Illustration Diagram**

IV. RELATED WORK

The presented guilt detection approach is related to the data provenance problem: tracing the lineage of *S* objects implies essentially the detection of the guilty agents. Suggested solutions are domain specific, such as lineage tracing for data warehouses, and assume some prior knowledge on the way a data view is created out of data sources. Our problem formulation with objects and sets is more general and simplifies lineage tracing, since we do not consider any data transformation from Ri sets to S.

As far as the allocation strategies are concerned, our work is mostly relevant to watermarking that is used as a means of establishing original ownership of distributed objects. Watermarks were initially used in images, video and audio data [2] whose digital representation includes considerable redundancy. Our approach and watermarking are similar in the sense of providing agents with some kind of receiver-identifying information. However, by its very nature, a watermark modifies the item being watermarked. If the object to be watermarked cannot be modified then a watermark cannot be inserted.

In such cases methods that attach watermarks to the distributed data are not applicable. Finally, there are also lots of other works on mechanisms that allow only authorized users to access sensitive data. Such approaches prevent in some sense data leakage by sharing information only with trusted parties. However, these policies are restrictive and may make it impossible to satisfy agents requests.

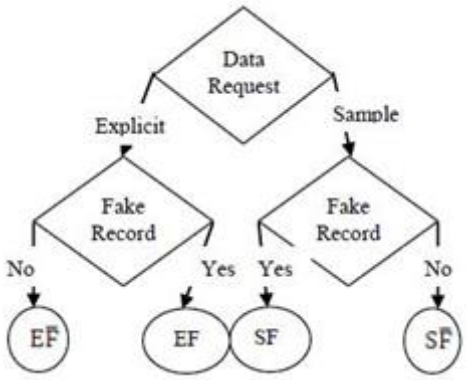

FIGURE 2: Leakage Problem Instances

CONCLUSION & FUTURE WORK

In doing a business there would be no need to hand over sensitive data to agents that may unknowingly or maliciously leak it. And even if we had to hand over sensitive data, in a perfect world we could watermark each object so that we could trace its origins with absolute certainty. However, in many cases we must indeed work with agents that may not be 100% trusted, and we may not be certain if a leaked object came from an agent or from some other source. In spite of these difficulties, we have shown it is possible to assess the likelihood that an agent is responsible for a leak, based on the overlap of his data with the leaked data and the data of other agents, and based on the probability that objects can be "guessed" by other means.

Our model is relatively simple, but we believe it captures the essential trade-offs. The algorithms we have presented implement a variety of data distribution strategies that can improve the distributor's chances of identifying a leaker. We have shown that distributing objects judiciously can make a significant difference in identifying guilty agents, especially in cases where there is large overlap in the data that agents must receive.

REFERENCES

[1]. Rudragouda G Patil, "Development of Data leakage Detection Using Data Allocation Strategies International Journal of Computer Applications in Engineering Sciences [VOL I, ISSUE II, JUNE 2011, [ISSN: 2231-4946].

[2]. S. Czerwinski, R. Fromm, and T. Hodes. Digital music distribution and audio watermarking.

[3]. L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression, 2002.

[4]. S. U. Nabar, B. Marthi, K. Kenthapadi, N. Mishra, and R. Motwani. Towards robustness in query auditing. In VLDB '06.

[5]. Stevens Le Blond, Chao Zhang Arnaud Legout, Keith Ross, Walid Dabbous, Exploiting P2P Communications to Invade Users' Privacy.

[6]. P. Buneman, S. Khanna, and W. C. Tan. Why and where: A characterization of data provenance. In J. V. den Bussche and V. Vianu, editors, Database Theory - ICDT 2001, 8th International Conference, London, UK, January 4-6, 2001, Proceedings, volume 1973 of Lecture Notes in Computer Science, pages 316-330. Springer, 2001.

