International Journal of Electrical Electronics & Computer Science Engineering
Volume 3, Issue 1 (February, 2016) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

Securing Data and Tracking Data Leakages in Business Applications

M. Vivekananda Swamy[1], Dr. M. Nagaratna[2]
[1]PG Scholar, Department of CSE, JNTUH College of Engineering, Kukatpally, Hyderabad, India
[2]Assistant Professor, Department of CSE, JNTUH College of Engineering, Kukatpally, Hyderabad, India
[1]vivekm.ktc@gmail.com,[2]mratnajntu@gmail.com

Abstract: ***Data security plays a vital role in small, medium and large scale business. Data experts understand the importance of data security and they try to secure data at every point. Today there is a great requirement for not only securing the data being shared but understanding its leakage points, as when it gets leak. After knowing the leakage point it becomes very essential to know who is the data leaker. While sharing data in business environment it becomes critical to understand the data leakers for the very first time to safeguard the business from great loss, during the leakage activity. In this paper, we implements secure sharing and data tracking mechanism between the two parties i.e. distributor and agents. In this scenario of data owner, agents and third parties whenever there is a leakage occurring to business data files wittingly or unwittingly to third party then the leaker of data file is tracked and detected exactly by the system.***

Keywords: Secure, Leakage, Detection, Data Files, Leaker, Third Party, Wittingly, Sharing.

I. INTRODUCTION

We are living in an information age where data plays a great role in performing major tasks of day today business activities. Data Leakage is nothing but unauthorized transmission of personal or sensitive data or information from an organization to unknown third party i.e., an unauthorized recipient. Today the businesses are so sophisticated dealing with their client data that one small leakage of data can take the business to shutdown or huge loss to the whole organization. The data in the business organization may be personal information(like credit card data, medical information), intellectual property and other information depending on the nature of the business or industry. Even though the information rapidly changes time to time, for the specific period the data has its own importance and it has to be secured and tracked during its movement from one party to another by some authentications. In any case the information leak during its movement to unauthorized third party then it become high priority to know how the leak has taken place and who is responsible for the information leakage, so that the guilty party can be traced, punished and also future data or information can be safeguarded well for the beneficial of the business.

Here in this case, we have much concentrated on not only securing the data transmission by providing the secured key but also tracking the business data or information by tracking objects which will always keep track of moving data files. When the data is leaked which is stored in database, then the leaker will be detected with very high probability. This approach gives a scenario where there is full chance of detecting the guilt party and securing future data leakages in a business environment.

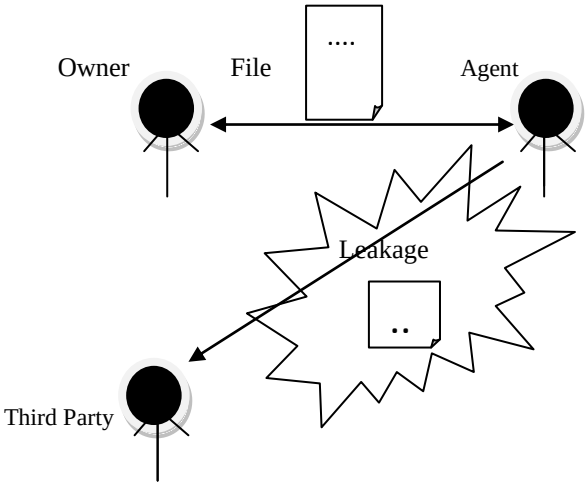

Fig.1 What Is Data Leakage?

The above figure explains the simple form of data leakage, where the business owner performing his task by giving the data files to his trust worthy agents but as time moves on the trust worthy agent turns into untrustworthy by leaking shared files of business owner to the third party. Once the business data goes into the hands of the third party without the notice of the owner, then it is termed as data leakage. Today social networking sites like Facebook, Twitter etc., along with their third party applications everyone using a part or whole of their users personal information which, they promise to keep undisclosed and secure. But there will be always a threat for the user personal data to leak and then it becomes necessary to catch the leaker and make sure the user data is always safe in future. The future of any user personal data always demands that it should be secured and have no leaks to unauthorized people who can miss use the data to any extent.

II. PROBLEM STATEMENT

The main problem focus is on data security and how to identify the data leaker has been analyzed in this paper.
1. Data security while sharing the data files in business is addressed by a secured key which is provided to the file receiving agent that is randomly generated and only the distributor and agent for whom the data file has

been sent will be knowing about the secured key. The agents who is in the system will only view the files using provided secured key. These file security keys are generated by the business owner and send along with the file to the agent to perform the desired business task on demand by the owner.

2. Identifying the data leaker can be explained as considering the problem as set of agents as P1,P2,P3.....Pn and a owner/distributor who is providing the data files to all these agents to perform his requested works and the work activity contains the inputs from the owner like the owner confidential data files, medical information or his client information, which is very sensitive and costly to the owner at that movement of time. If any leak occur here there will be a great loss to the owner business. Here the agents P1,P2,P3.....Pn take the data from the owner and will do the work as per the request from the owner. If the agent is sincere then there will be no problem at all, the designed technique is mainly used in the situation to catch hold of the data leaking agent with very high probability. The data leakage process goes on in the following way

• Owner gives the data to the agent whom owner has made the contract to perform his activities.

• Agents takes the data files as inputs and performs the work of the owner and submits to him.

• Whenever the sensitive data of the owner is given to third party by the agent then the problem arises and the data leakage will take place.

• When agents leaks the data of the owner then the data tracking objects inserted to monitor the malicious activity of the agent, will note leaker information who has performing the mischief activity will be recorded and handover to the business owner.

• As soon as the business owner receives the information he takes the action towards the culprit/leaker agent to save the future business data.

III. RELATED WORK

Early work in the area of data leakage detection resulted in the idea of using watermarks within sensitive digital information.[1] Here, a uniquely identifying text or image is embedded within each copy that is distributed to authorized agents. When leakage occurred, then this unique code would help identify the party that was responsible for the leak. The problem with this approach was that even though this is an easy solution, it still involves a certain modification of the original data information set. Also, it was observed that such watermarks could be tampered with to Sufficiently distort the uniquely identifying code or sometimes completely destroyed if the data recipient is malicious. Papadimitriou and Garcia-Molina [2], two Stanford researchers proposed a non-obtrusive leakage detection system which can detect the guilty leaker without changing the integrity of the original data. In their paper, they proposed the main premises on which much work in this area has been based. They propose several data allocation strategies (across the agents) that improve the probability of identifying leakages. The chief contribution here was that the proposed techniques were not based on altering the distributed data in any way, but allocating the data intelligently so as to identify the guilty party. In their work, Agarwal and Gaikwad [3] dealt with data leakage issues that arise from popular applications like email, IM and other Internet channels. E- Mail filtering was dealt on the basis of the fingerprints of message bodies, the white and black lists of email addresses and the words specific to spam. Also, in the case of data leakage from trusted agents, the distributor must evaluate the odds that the leaked records came from one or more agents. For this purpose, they used data allocation strategies or injecting "realistic but fake" data records to improve identification of leakage. Jagtap et al. **[**4] implemented a system called the Data Watcher and Leakage Detector to detect and prevent data leakage. The authors developed two models - first, if data leakage occurs when an employee of an enterprise accesses confidential data without the consent of owner, the Data Watcher model is used to identify the data leaker. Second, if data leakage occurs when an employee has given data outside the enterprise, then a second model called the Leakage Detector is used for assessing the “guilt” of the involved parties. Their Guilt model uses fake objects as a watermarking tool to improve the probability of identifying guilty third parties.

Ajay Kumar, Ankit Goyal, Ashwani Kumar, Navennet .Kumar Chaudhary and Sowmya Kamath S, [5] "Comparative Evaluation of Algorithms for Effective Data leakage Detection" In this the author handles the data leakage by comparing various algorithms techniques like round robin, fcfs, srf, lrf for effective data leakage detection and also says that round robin algorithm has very high probability of finding the guilt agent.

IV. SYSTEM DEVELOPMENT

In this part , we will discuss things like how the system will share data files among the work making agents inside the business and if any agent give the data to other agents or third party i.e leaking data then the owner will come to know the data leakage has been done and which agent has done the leakage. In this system the tracking mechanism of data leakage uses the concepts of tacking objects for monitoring the spurious activities carried out in the whole developed system. All the activities are well monitored while sharing the data from source to destination. Tracking objects which are at agent side will continuously watch for agent actions, if in any case the agent is involved in the file data

International Journal of Electrical Electronics & Computer Science Engineering
Volume 3, Issue 1 (February, 2016) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

leakage activity then the system triggers the information that the leakage has been done and say exactly the name of the agent who has performed the data leakage.

Secure sharing of data: The data is shared to all the desired agents with a randomly generated security key, only by which the agent can access and view the file sent by the distributor. Distributor can view any file at any time but the agents must have the key to view the current file sent by distributor.

Tracking objects for monitoring: These objects are main monitoring objects by which the distributor will come to know the malicious activities performed by various agents. As these objects are named tracking objects, they will keep tracks of all the agents actions in the system and ready to report to the distributor.

Fig2 explains the typical process of how the whole system is developed. In a taken business model, there is a relationship between one business owner and many agents.

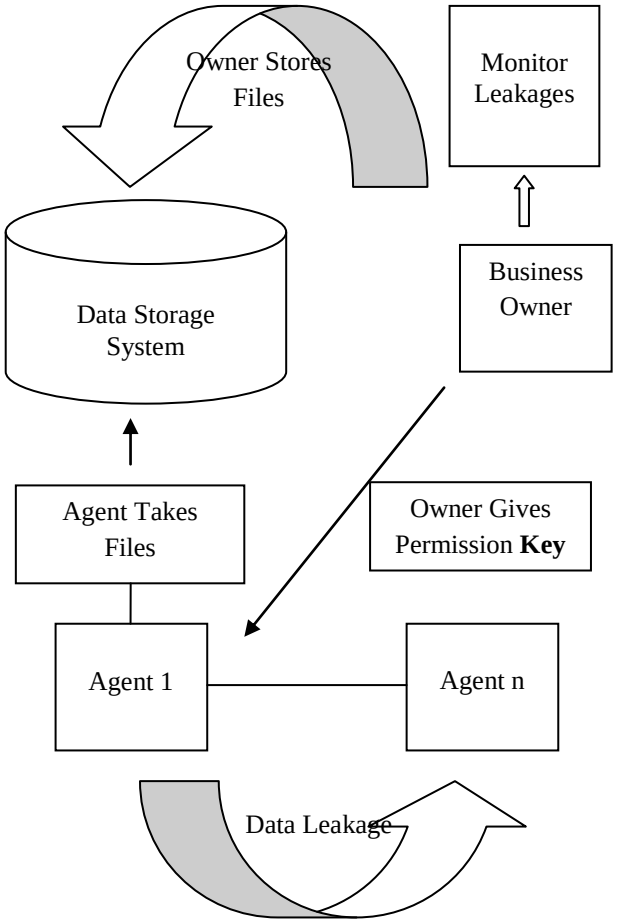

Fig.2 Scenario of Data Leakage in Business

Business owner stores the data files in database and provides the access permissions to the agent like as shown agent1 is getting permission key from business owner. Once the agent gets the permission he can access the file from the databases directly. The data files can be leaked between agent 1 to agent n or vice versa without the notice of business owner. Through the monitor unit the business owner will track all the activities of the agents available in the business. Data Allocation strategy used in this system is random allocation to the agents from the distributor. All the agents say x agents will be registering in the system and available for the distributor for taking task and completing it. Then the distributor decides and choose the agent to whom the data has to be allocated on random basis where for future reference he has the monitoring facility that tells who has been allocated which data.

Some special features of this system is that it is developed in object oriented language and it contain two user logins one is owner and the other is registered agent.

In the following system, both the users perform their roles based task by data sharing to one another, in order to meet the business needs. When the task making user i.e agent leak the data to the third party, which is considered as a threat to the business then the business owner who is providing the data file will come to know by the tracking and detecting system by his login into the system.

A. Safe State Activity Algorithm:

- Distributor D selects agent a_1,a_2,a_3..... a_n.
- Send file F to agent a3 with secure key say X.
- Agent a3 receives the alert message and key.
- Agent a3 using key X, access the file F.
- Agent perform the given task and submit the results to distributors.

B. Unsafe State Activity Algorithm:

- Distributor D selects agent a_1,a_2,a_3..... a_n.
- Send file F to agent a3 with secure key say X.
- Agent a3 receives the alert message and key.
- Agent a3 using key X, access the file F.
- After the file access, current Agent performs the malicious activity of leaking the file F to third party.
- Tracking objects o_1,o_2,o_3... gets activated and find the malicious activity performed by agent.
- Leaker information is send to the distributor D.
- Distributor will see the information of leaking agent and take the action.

V. RESULTS

Event based detection is seen in this whole process of this project. For every event of file sending from distributor to agent there will be controlled detection of agent actions. If the agent performs any kind of mischief activity then leakage value is obtained from

International Journal of Electrical Electronics & Computer Science Engineering
Volume 3, Issue 1 (February, 2016) | E-ISSN : 2348-2273 | P-ISSN : 2454-1222
Available Online at www.ijeecse.com

leakage monitoring objects. This value is known to be leakage detection value(Ldv) and the Ldv is zero for no leak cases and one for the leak cases.

The formula obtained for leakage detection on event value is written as delta equals to leakage detection value.

$$\Delta = 1 + ldv$$

where, 1 is the file/event constant and ldv is leakage detection value. Graphically, the application is tested for 10 events and the following graph obtained

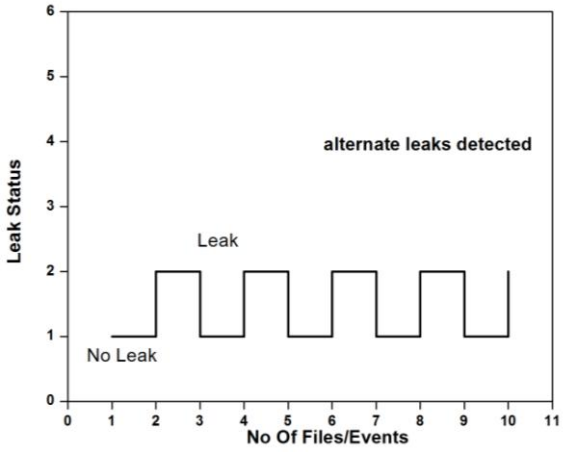

Fig.3 Leakage Detection of 10 Files/Events

The above graph specify the real leak status of ten events. Here the event is file sent to agent and the action the agent performs. The above plot show the alternate files has been leaked by various agents where we see sudden phase shift. The digital signal in the Fig.3 specifies that for every alternate file sent to the agent, there is a leak observed by that particular agent. Hence that particular agents are reponsible for the file leakage in the system and they are caught and named as guilt agents.

VI. CONCLUSIONS

In this work, we concluded that any authorized user who is receiving the data or information can leak the data to third party by using his authorized login credentials. When he leaks the data wittingly or unwittingly the information of leak is sent to the owner or distributor by the background tracking sql objects. These objects are embedded in the soft code and they track the behavior of the users who are performing the task for the owner. We have much concentrated on tracking the business data or information by storing the data in database and when the data is leaked the leaker will be detected with very high probability. This approach gives a scenario where there is full chance of detecting the guilt party.

The future work which can be derived from this project may be data leakage prevention. Where in this paper data security is tried by the key supply to the working agent without which the file cannot be shared and detection of data leaker if any untrustworthy agent is doing the malicious activity. The next important and essential thing will be prevention of sensitive data leakage when the malicious activity is tried. One more thing that can be extended is agent data allocation on request through online mode.

VII. REFERENCES

[1] R. Agrawal and J. Kiernan, “Watermarking Relational Databases,” Proc. 28th Int’l Conf. Very Large Data Bases (VLDB ’02).

[2] P. Papadimitriou and H. Garcia-Molina, “Data Leakage Detection,” technical report, Stanford Univ., 2008.

[3] Ankit Agarwal, Mayur Gaikwad, Kapil Garg, Vahid Inamdar, “Robust Data leakage and Email Filtering System”, International Conference on Computing, Electronics and Electrical Technologies, 2012.

[4] N. P. Jagtap, S. J. Patil, A. K. Bhavsar, “Implementation of data watcher in data leakage detection system”, International Journal of Computer & Technology Volume 3,No. 1, Aug, 2012.

[5] Ajay Kumar, Ankit Goyal, Ashwani Kumar, Navennet Kumar Chaudhary and Sowmya Kamath S, " Comparative Evaluation of Algorithms for Effective Data leakage Detection."

[6] Sridhar Gade, Kiran Kumar Munde and Krishnaiah.R.V. "Data Allocation Strategies for Leakage Detection."

[7] Ankit Agarwal, Mayur Gaikwad, Kapil Garg, Vahid Inamdar, “Robust Data leakage and Email Filtering System”, International Conference on Computing, Electronics and Electrical Technologies, 2012.

[8] Y. Cui and J. Widom (2001) ‘Lineage Tracing For General Data Warehouse Transformations’ -In the VLDB Journal,pp. 471– 480.

[9] Preeti Patil, Nitin Chavan, Srikantha Rao, S B Patil, “Building of a Secure Data Warehouse by Enhancing the ETL Processes for Data Leakage”, Intl Conf & Workshop on Recent Trends in Technology,2012.