RESEARCH ARTICLE OPEN ACCESS

Data Leakage Detection and Prevention System

Chirag S. Patil [1] , Swapnil S. Nalawade [2], Vikas D. Natekar [3], Prof. Neha Saxena [4]
Bachelor of Engineering
Department of Computer Science and Engineering
Theem College of Engineering
Boisar (east), District Palghar
Maharashtra – India

ABSTRACT

In previous technologies, threat in cyber-attacks are increasing because existing security system and mechanism are not able to detect them. Past cyber-attacks had simple purposes of leaking private information by attacking the Personal computer or destroying the system. However, the goal of recent hacking attacks has changed from leaking information and destruction of services to attacking large-scale systems such as critical infrastructures and agencies. Nowadays internet technologies has become the main factor of any business organization. These organizations use this technologies to improve their efficiency by sending data from one location to another. But, there are number of threats in sending critical organizational data as any culprit employee may public this data. This problem is known as data leakage problem. In this proposed work, we are suggesting a model for data leakage detection and prevention. In this model, our aim is to identify the culprit who has leaked the organizational data. Identify data leakages from distributed data using some data allocation strategies and find out the agent who leak that data. Improve probability of finding out guilty agent and provide the security to that data.

Keywords:-Message Chaining, Intrusion Detection, Fake Object, Guilty User, Agent, Distributor.

I. INTRODUCTION

In the running business scenario, data leakage is a big challenge as critical organizational data should be protected from unauthorized access. Data leakage may be defined as the accidental or intentional distribution of private or important organizational data to the unauthorized entities. Its very important to protect the critical or private data from being misused by any unauthorized use. Critical or private data include intellectual copy right information, patent information, functional information etc.

In many organizations or agencies, this private organizational data have been shared to outside the organizational premises. Hence, it is difficult to identify the culprit, who has leaked the data. In the proposed work, our goal is to identify the guilty or culprit user when the organizational data have been leaked by some agent. In the proposed work, security model has been used which provide the analysis and design of secure computer systems. According to the report, sophisticated hacking attacks are more increasing in the cyber space. Hacking in the past leaked private information, but latest hacking targets companies, organization, government agencies. This kind of attack is called APT (Advanced Persistent Threat). APT targets on a specific system and analyses vulnerabilities within the system for a long time. Therefore it is very hard to prevent and identify APT than traditional attacks and could result system damage. detection and prevention systems for defending against cyber-attacks were intrusion detection systems, firewalls, intrusion prevention systems, database encryption, anti-viruses solutions, DRM solutions and etc. more than, integrated detecting technologies for managing system logs were used. These security solutions are design based on signatures and blacklist. However, according to various research, IDS and prevention systems are not capable of protecting systems against such type of attacks because there are no signatures. Therefore to overcome this issue, security communities are beginning to apply data mining and various technologies to detect previously. unknown attacks. For example, web services, such as email for communicating with others either within or outside of an organization, but introduce the risk of data transformation. Intrusion Detection Systems (IDS) as well as Intrusion Prevention Systems (IPS) are frequently used to protect networks from cyber attacks. In particular, such systems can prevent confidential information by blocking accidental or intentional leakage.

To provide such functionality, these systems based on client-server inspection they allowed the definition of configuration rules. While today's IDS and IPS and prevention systems perform well for unencrypted traffic, they struggle with encrypted traffic, resulting in poor

performance. As a workaround, the secure and encrypted channel from the Internet is often terminated, which mounts some kind of man-in-the-middle-attack. This solution ensures an effective detection and prevention. In intrusion detection may not only be desirable and relevant in the text of enterprise networks, but is also gaining in importance in today's trend to outsource the network management, including security to third parties. For e.g, the management of third-party networks can be a lucrative business for Internet Service Providers. At the same time, for customers running security critical businesses for e.g banks, it is important that the privacy of traffic be preserved. We in this paper however observe another confidentiality issue of today solutions: it concerns the confidentiality and integrity of the inspection logic itself. For example, the development and maintenance of effective IDS rules is challenging, and especially small company do not have the expertise and time to define the most effective rules and constantly follow the new. It constitutes a business opportunity for third parties: a company specialized into security search can take over the responsibility to define and follow a good rule for company. However, a business model also introduces new requirements. In a third party company or agencies may not be willing to share its rules.

The rest of the paper is organized as follows. We review Literature Survey in Section II. In Sections III, we present Proposed System. Section IV Architecture. In Section V, Working. We give conclusion in Section VI.

II. LITERATURE SURVEY

A. Comparative Evaluation of Algorithms for Effective Data Leakage Detection:

In the real time networking, unknown and intelligent threats are increasing. These unknown attacks cannot be detect or mitigated using old pattern matching methods such as rule, and black list based solutions. Traditionally, data leakage detection is handled by watermarking, e.g., a unique code is embedded in each copy. If that copy is later found in the hands of an unauthorized party or outside agencies, the leakier can be identified. Watermarks can be very useful in some cases, but again, involve some modification of the original data. However, watermarks can sometimes be destroyed if the data recipient is malicious. The Existing System can detect the hackers but the total no of evidence will be less and the organization or authorized person may not be able to proceed legally for further proceedings due to lack of evidence and the chances to escape of hackers are more.

Disadvantage of the system:

- Hacker can delete the watermark.

B. **Detecting data semantic: A data leakage prevention approach :**

In this paper, the effectiveness of using analysis techniques to detect confidential data semantics was studied. A DLP classification model was proposed based on the well-known information retrieval function. The classification was based on measuring the similarity between the documents and the category. This model was tested against different scenarios with known, unknown data, and partially known,. The overall classification shows and encouraging outcomes across all scenarios. Further, a graphical representation of the classification results was applied using abstraction. The visualization provided a very useful analytical tool for studying the semantics of documents in relation to category centroids. These results shows 60 percent of the modified documents were able to be identified.

Disadvantage of the system:

- Only 60 percent of the modified documents were able to identified.

C. **Data Leakage Prevention System with Time Stamp:**

In this paper, document along with the data one more parameter i.e. time stamp also considered as an important aspect in the Data leakage Prevention. For example in an educational system question paper is confidential until on or before examination date once exam over that is public and treated as a non-confidential. In Learning Phase the documents are trained a confidential documents with time stamp. In the detection phase the tested document is compared with confidential score and time stamp, if the time stamp of the tested document is greater than or equal to the time stamp in the above table then that document is treated as a confidential and it is blocked.

D. **Privacy Preserving Inspection of Encrypted Network Traffic**

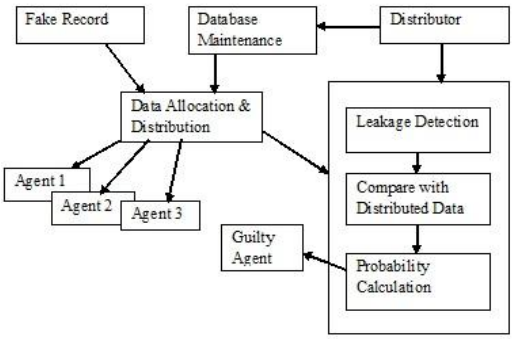
This paper, they develop the classic problem of traffic inspection from an interesting new, privacy preserving perspective. while today it is commonly believed that it is inevitable that users have to blindly trust the administrator managing the intrusion detection or prevention system, in this paper, questioned this assumption. In particular, they shown that it actually is possible to reduce trust assumptions in the enterprise network, and presented an intrusion detection system which is not only privacy preserving regarding the user traffic but also regarding the rules used in the IDS/IPS. the proposed PRI system requires a single secured

server; no modifications of the hardware at the users is required.

III. PROPOSED SYSTEM

In the proposed system we study techniques for detecting leakage of a set of objects or records. Specifically, we study the following scenario: After giving a set of data to agents, the distributor discovers some of those data in an outside or unauthorized place. For e.g, the data may be found on a website, or may be obtained through a various legal discovery process. At this point, the distributor can assess the possible that the leaked data came from one or more agents, as opposed to having been independently gathered by other means. In the proposed system, we develop a system for identifying the "guilt" of agents. We also present algorithms for distributing objects to different agents, in a way that improves our chances of identifying a data leaker. Finally, we also consider the option of adding "fake" objects into the distributed set of data given by distributor. Such objects do not correspond to real entities but appear realistic to the agents for this we add fake object. In a sense, the fake objects act as a type of watermark for the given entire set, without modifying any individual agents. If it find that an agent was given one or more fake objects that were leaked, then the distributor of system can be more confident that agent was guilty. In the Proposed System the data leaker can be traced with good amount of evidence.

IV. ARCHITECTURE



"Figure 1: Basic architecture of the system."

In recent year where more services and resources are outsourced, there is an increasing need for maintaining solutions for security. Over the few years, this problem has been increase in various company or infrastructure may expect their data to remain confidential. we develop system for identifying data leaker from any agencies. A distributor can insert original as well as fake records in the Database. A agent can be registered or login by entering details such as name, email id, password. A registered agent can Login and make a request to the distributor for data. distributor can add fake object into the data ,which is required by agent and send it to the agent. The system then provide the requested data from the database and performs the operation of fake records to the set of original records. Then it provides this data to the agent. Without distributor permission agent cannot send this data to outside of organization. If agent may pass on this data to an unauthorized party then the file which is send by agent will be send as bank file. Using leakage detection and fake object distributor identified guilty agent and its details. The below figure shows the basic working of the proposed intranet system for student's attendance management system. The below figure shows the basic working of the proposed intranet system for student's attendance management system.

V. WORKING

Our system have Customized software for file shearing and have its own file format. While shearing the file with employee within organization it will ask for admin permission if admin grant the permission for shearing employee will shear the file. Login page for all users will be shown, such as agent and distributor. Once user login, according to their login details (i.e. username and password), the user will be identified that whether the user is Distributor or agent.

Once user identified the respective JSP page will be shown where the distributor can send the files to agents as well as monitor the file sharing among the agents. While sending the file from distributor to agents some fake objects to the file which is being sent to agent. If any agent is sending the same file to any other agent it will be identified and marked him/her as guilty Distributor can identify the guilty agents while monitoring the agents. File content will be deleted except hidden object if any agent is sending file to another agent.

VI. CONCLUSION

In this, we study the possibility that an agent may be responsible for data leakage using some techniques. We analyzed that distributing data may improve the chances of detecting the agents effectively specially when there is a large overlap in the data that agents must receive. Our objective was to verify the results of algorithms that finding the guilt agent among company. Hence, we can conclude that if the distributor wants to completely satisfy an agent before allocating any object to other agents our technique must be used in order to improve the chances of identifying the leaker. Our future work

includes the implementation of data allocation strategies for explicit data requests. We will also extend our work to handle agents' requests in an online fashion, i.e. when the number of agents and agent requests are not known in advance.

REFERENCES

[1] Ajay Kumar, Ankit Goyal, Ashwini Kumar, Navaneet Kumar Chaudhary, Sowmya Kamath S, "Comparative Evaluation of Algorithms for Effective Data Leakage Detection".

[2] Sultan Alneyadi, Elankayer Sithirasenan, Vallipuram Muthukkumarasamy "Detecting data semantic: A data leakage prevention approach".

[3] Subhashini Peniti, B.Padmaja Rani "Data Leakage Prevention System with Time Stamp".

[4] Liron Schiff, Stefan Schmid "Privacy Preserving Inspection of Encrypted Network Traffic".

[5] Marco Pistoia, Omer Tripp, Paolina Centonze, Joseph W. Ligman "Labyrinth: Visually Configurable Data-leakage Detection in Mobile Applications".

[6] Jos´e Ortiz-Ubarri, Eric Santos, Humberto Ortiz-Zuazaga "A Web Based Network Flow Data Monitoring System at Scale".

[7] Panagiotis Papadimitriou, Hector Garcia-Molina "A Model for Data Leakage Detection".