IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10432

Data Leakage Detection and Prevention for Cloud Environment

Mrs Rohini B. Gurav[1], Priya A. Kunwar [2], Vaibhavi R. Pawar[3], Parinita R. Waghmare[4]

Lecturer, IT,AISSMS's Polytechnic, Pune, Maharashtra, India[1]
Student,IT,AISSMS's Polytechnic, Pune, Maharashtra, India[2]
Student,IT,AISSMS's Polytechnic, Pune, Maharashtra, India[3]
Student,IT,AISSMS's Polytechnic, Pune, Maharashtra, India[4]

Abstract: Nowadays, many organizations do the business form one level to another level. They do the business through the different agents and organizations. With the increment of the business level the storage and transmission of the data via one place to other also increased so in the middle may be any user can leak the data. Our main focus is to detect the guilt agent who has leak the Sensitive data. Information leakage can be referred to as the unofficial move of arranging data from a server area to the outside world. In those papers, it describes the appropriation model for data leakage counteractive action and chooses a document allotment plan with minimal cover between the sets of records of the clients, as a result, it can detect leaked sources with high probability. Sensitive information in organizations may be company internal policies, money related data, individual charge card information, and organization-related data, this type of sensitive data can be a leak by a malicious user. Information leakage leads to a major issue for various organizations. In some Data Leakage Detection model used the 'fake objects' which are store in the server database. The fake objects help to identify the user who has leaked the file. Every user has a probability to leak the file which probability is called the guilt probability. Those users who have the probability to leak the file is known as the guilt probability. Many models of detection of the data leakage focus on the fake object which is including with the database to find out the leaker.

Keywords: data upload, data download, unauthorized access, data leakage, malicious user, Role base access control.

I. INTRODUCTION

Many companies nowadays conduct business from one stage to the next. They carry out their business by various agents and organisations. When the business level rises, so does the collection and transfer of data from one location to another, making it possible for any user in the centre to leak data. Our main goal is to find the guilty person who leaked the sensitive information. The unofficial movement of data from a server area to the outside world is known as information leakage. The appropriation model for data leakage counteractive action is defined in those articles, and it chooses a document allotment plan with minimal cover between the sets of records of the clients, as a result, it can detect leaked sources with a high probability. Business internal rules, money-related data, individual charge card information, and organization-related data are examples of confidential information that can be leaked by a malicious person. Information leakage is a big problem for a variety of organisations. Some Data Leakage Detection models make use of "false objects" stored in the server database. The fictitious objects aid in identifying the individual who leaked the file. Any consumer has a chance of leaking the file, which is known as the guilt likelihood. The guilt likelihood refers to the number of users who have a chance of leaking the file. Many data leakage detection models concentrate on the fake entity that is included with the database in order to identify the leaker.

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,

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10432

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 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.

II. LITERATURE SURVEY

The distributor allocates the data to an agent with a constraint and objective. The constraint is to fulfil the agent request for the file. The distributor maintains the table for the user request. The objective can detect the leak. The distributor chooses the minimum path to send their file to the user who is requesting the particular file. The model used the minimum transaction of a file to fulfill the client user. Suppose any file transfer A source to B destination in between through only two mediators which are X and Y, if the file is leaked while the transaction process then the probability of leaked file is (½). The leaker of the file can be X or Y. But if we do not use the minimum overlap method and the same file transfer from A source to B destination in between there can be N number of the mediator (n1, n2, …, nN). If the file is leaked during the transaction process then the probability becomes (1/n), which is something more difficult to find malicious users. So, minimum overlap is an efficient method to find the guilt user [1].

Fake object is created by the distributor. The distributer added some fake object with the actual data to improve the efficiency of detecting the leakage source. The object is not real. It's only generated at the transaction time to trace the guilt agent that is supposed to leak the data. The fake object is generated in a way where the agent can't distinguish between a fake object and a real entity. Fake records generally used to actual data for tracking or monitoring their data for example: X company sell an item to company Y in which company X add some fake tracing software include their company address if company Y misuse their data to sell outsiders their data then automatically receive the copy of issues records so that company X can easily identify the improper use of data. The fake is denoted by the F= {f1, f2,…, fn} [2].

In this paper by S. N. holambe [3], the distributors detect the agents by sample and explicit algorithm. In the sample algorithm, it depends on the fake object which is put on the original data to improves the chance of guilt detection the only difference is the guilt user only detects when he received the fake leaked objects but in the explicit request, the fake objects are not followed its all only depends on the agent request. The distributor to add some static fake objects [3].

If the database is corrupted, then chances arise to lose the data. Somehow it is possible to recover the lost data from a corrupt file with the right applications. If you are working on any electrical devices such as laptops, PC, etc. Suddenly power off while working then have the maximum chance to lose their confidential data. So, avoid this problem by saving your work regularly [4].

The data or file leaked or loss by the unauthorized clients or user's that denoted by the L. Since clients having some of the leaked data of L may be high strung for leaking the information or data. After the leakage of the file, he may argue that he is innocent and that the L data were accessed from the distributor and send the target users or target organization via some other way. Our aim to find out the users or clients who leak the file most. For example, if any of the objects or data of L was sent to only agent A1, we may only suspect A1 more because only A1 users access the file set. So that probability of that client A1is guilt leak probability (G1) dataset L is defined by the formula as{G1|L} [5].

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 4, April 2021

DOI 10.17148/IJARCCE.2021.10432

III. CONCLUSION

From this study we conclude that Data leakage is a silent but destructive type of threat.Sensitive information can be leaked without any knowledge.It may be an insider work or may be an outsider.Sensitive data must be secured by security model has been used which provide the analysis and design of secure computer systems, security communities are beginning to apply data mining and various technologies to detect previously. unknown attacks, Intrusion Detection Systems (IDS) as well as Intrusion Prevention Systems (IPS) are frequently used to protect networks from cyber attacks.The data leakage detection system model is very useful as compare to the existing watermarking model. We can provide security to our data during its distribution or transmission and even we can detect if that gets leaked. Thus, using this model security as well as tracking system is developed. Watermarking can just provide security using various algorithms through encryption, whereas this model provides security plus detection technique.Our model is relatively simple, but we believe that 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. Our future work includes the investigation of agent guilt models that capture leakage scenarios.

IV. REFERENCES

[1] Yin Fan, Wang Lina, Yu Rongwei, Ma Xiaoyan"A Distribution model for Data Leakage Prevention," 2013 IEEE International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC 2013), Shenyang, China.

[2] Yadav Gitanjali B., Bhaskar P. C., Kamat R.K, "Assessing the Guilt Probability in Intentional Data Leakage," 2012 International Journal of Computer Science and Information Technologies.

[3] Sushilkumar N. Holambe, Ulhas B. Shinde, Archana U. Bhosale "The Guilt Detection Approach in Data Leakage Detection," 2015 International Journal of Computer Applications.

[4] Dr. A R. Pon Periyasamy, E. Thenmozhi "DataLeakage Detection and Data Prevention Using Algorithm" (2017) International Journal of Advanced Research in Computer Science and Software Engineering.

[5]S. Praveen Kumar, Y. Srinivas, D. Suba Rao, Ashish Kumar, "A Novel Model for Data Leakage Detection and Prevention in Distributed Environment," 2016 International Journal of Engineering and Technical Research (IJETR).