International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 2 | ISSN : 2456-3307

Data Leakage Detection Using Client-Server Mechanism

Harshala Khapekar[1], Sanskruti Gunde[2], Swati Shingne[3], Pooja Shende[4], Drishti Moon[5]

[1, 2,3,4,5] Department of Computer Science & Engineering, Rajiv Gandhi College of Engineering & Research, Nagpur, Maharashtra, India

ABSTRACT

Data Leakage is defined as an unintentional distribution of sensitive data to an unauthorized agent or entity. Internet act as the main backbone in business organizations and in every other prospect of life. Though internet plays a significant role still there are certain numbers of security issues in transferring the sensitive data as any culprit employee may public this sensitive data. A data distributor has given sensitive data to a set of his trusted employees. Some of the data is leaked and found in an unauthorized place. The distributor needs to evaluate and find out that leak data came from which culprit employee or the unauthenticated agent. This model proposes data allocation strategies to improve the possibility of identifying the leakage. These methods do not depend upon the alterations of the released data. This model introduces realistic but fake data objects to further improve our chances of detecting leakage and identifying the culprit. Finally, this mechanism shall be then implemented on the cloud server.

Keywords: Data Leakage, Data Allocation Strategies, Fake Objects, Perturbation

I. INTRODUCTION

Security in a particular network or in any cloud server states that the data or information which is available on the cloud server can only be accessed or viewed by authorized users and it must prevent the data or information from any unauthorized user. Administrator of a particular network is the person who provides authorization for users to access their data. If the cloud server is private then security should be provided within an organization or a company whereas, in a public cloud security should be provided globally by the use of username and password [1]. Only the person who has authorization can use it and others are restricted. In some cases where business is carried out, sometimes sensitive data must be handed over to supposedly trusted third parties [3] [4]. Our goal is to detect the data leakage when the sensitive data of the data owner has been leaked by his one of the agent or more, and if possible to identify the agent that leaked the data. This model considers applications where the original data cannot be perturbed. Perturbation is very useful technique where the data is modified and made “less sensitive” before being handed to the agents. However, in some cases it is not feasible to change the original distributor’s data e.g. if an outsourcer is paying a salary to an entity, he must have the exact salary and customer bank account number details. If medical researchers will be treating patients they may need accurate data for the data. Data leakage [11] may be defined as an accidental or intentional distribution of private organizational data to the unauthenticated agents. It is important to protect the sensitive data from being misused by any unauthorized user. Sensitive data may include intellectual copyright information, patent information, financial information or any kind of information of the organization, etc.

II. LITERATURE REVIEW

Traditionally, watermarking techniques were used for handling data leakage detection and it can be used to transfer the secured message from one destination to another destination. Watermarking has an objective to identify a data owner and hence is subject to attacks where a plagiarized person claims about ownership of that data [5]. This technique use the concept of message authentication and ensure that any change in message can be easily traced out during active attacks like masquerade, replay, modification of data, etc but it fails while in case of passive attack like traffic analysis and release of message contents. In watermarking, a unique code is embedded in the original data copy and if that copy of data is later found with an unauthorized person maliciously then leaker can be identified[6][9].The main disadvantage is that it requires some modification of the data which is termed as perturbation. At some conditions, watermarks can be destroyed if the data receiver is malicious. Hence, there is need to propose much more efficient technique to find out the data leakage.

III. METHODOLOGY

In proposed system, the given model removes the disadvantages of watermarking by adding fake objects thus increasing competence of the system. Adding “fake” object shall not correspond to real data but appears to be the same. If it turns out that an agent was given one or more fake objects that were leaked, then the distributor can be confident that agent was a fraudulent one[12]. This model develops a system for finding the fraud employees or unauthenticated agents and for this purpose, different data allocation strategies are used. So, distributor’s sensitive data remains safe.

A. Data Allocation Strategy

The main prospect of the project is the data allocation strategy which states the mechanism of how sharply distributor distributes his data to his employees to maximize the chances of detecting the culprit agent [7]. Here, Administrator can send the files to his authenticated users by sending a secret key on the user’s mail. User shall view these secret key details and use it to download the original data. If an unauthenticated person attempts to view it maliciously with a wrong key then related fake data object will be downloaded by him and hence, the owner can detect the leakage by maintaining the details of numbers of fake objects downloaded.

B. Fake Object

Fake objects are the data objects that are generated by the distributor or by the owner in order to increase the chances of detecting those agents who leaked the data [8]. The data owner may be able to add fake objects to the distributed data in order to improve his efficacy in detecting guilty culprits. In this case, if given the wrong secret key to download the file, the duplicate file or the fake object file will be opened and the fake object details will be sent to the data owner [10]. Here initially there is application of a certain kind of cryptographic algorithm to the text and then embedding the result to the document this process will guarantee the security of the data. An effective light weighted algorithm like AES can be used for the encryption.

IV. PROPOSED METHOD

The brief idea of the model can be briefly explained with the help of the block diagram given in the fig.1:

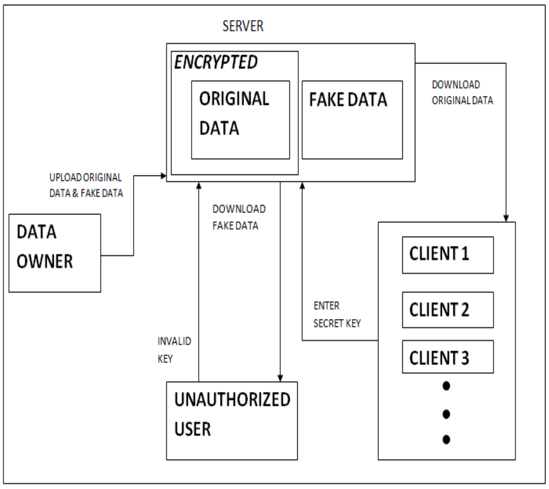

Figure 1. Block Diagram using client-server architecture

A data owner uploads the original data on the server as well as its correspondent fake data. The original data is in encrypted form but the fake data is not. When the authenticated employee enters the valid secret key he is able to download the original data. In case if the wrong key is entered the fake data will be downloaded.

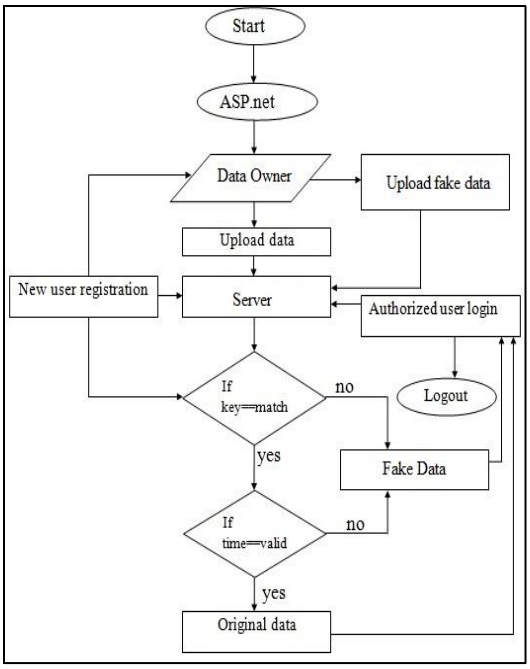

Figure 2. Flowchart

A. Security

This model ensures the security by using a secret key. Following is the algorithm for the generation of the key.

Algorithm for key generation:-

1) Compute the file size.
2) Extract random character from file name.
3) Merge the file size with extracted character from file name and special character(SC).
   Key= merge (FS, FN, SC)
4) Generate secret key using random function.
   Secret key= Random (Key)

V.RESULT

The generation of the valid secret key is done by using the proposed algorithm. Size of the file is computed initially and converted into bytes. Tabular representation of generation of the key is given below:

Table 1. Key Generation

Parameters				
Sr. No.	Size of File(FS)	Letters of File Name(FN)	Special characters	Key Generated
1.	20KB	Packagelist.pdf	!@#$	!pa$OPACKCK@#967
2.	30KB	ElectiveIII_CCC.pdf	!@#$	IC@#$03LE!ee96EC
3.	17KB	Mobile_main.pdf	!@#$	b@#31MBI!moi$490
4.	25KB	Information_Cyber.pdf	!@#$	!info#$01INO@18F
5.	21KB	Distributed_OS.pdf	!@#$	!dis#$1231ITt@DS

Name of the file to be uploaded is considered and random letters are extracted from it. Further special symbols are used and finally the letters, symbols and size of the file together are responsible for the generation of key.

The details of the leaked data will be sent to the administrator, if the wrong key is entered by the user or any unauthenticated user tries to leak it. The details of the leaked data are given below:

Table 2. Details of data leaked

File_name	Mob. No.	User name
Block diagram.png	7709053898	Sanskruti
Waterfall_model.jpg	9156593518	swati

VI. CONCLUSION

The proposed model ensures the data allocation strategy along with the fake object data so as to improve the chances of data leakage detection and ensure the security of confidential data. The fake objects that are proposed in this model are not realistic but they appear to be the same.

In the future work, the proposed model can be extended by generating dynamic fake objects as per the clients request and advanced security algorithms can be implemented to improve the chances of detecting the culprit1.

II. REFERENCES

[1]. V. Vijaylaxmi, T.Rohini, S.Sujata, A.Vaishali, "Survey on Detecting Leakage of Sensitive Data", IEEE,2016.

[2]. Neeraj Kumar, Vijay Katta, Hitanshu Mishra, Hitendra Garg, "Detection of Data Leakage in cloud computing Environment", IEEE,2014.

[3]. Panagiotis Papadimitriou, Hector Garcia-Molina, "Data Leakage Detection", IEEE, Vol-23, No.1, January 2011.

[4]. Rekha Jadhav, "Data Leakage Detection", International Journal of Computer Science and Communication Network", Vol(3),37-45.

[5]. Nikhil Chaware,Prachi Bapat,Rituja Kad,Archana Jadhav,Prof. S.M.Sangve, "Data Leakage Detection", Journal of Information Knowledge and Research in Computer Engineering,Volume-2,Issue-2,Oct 13.

[6]. Mr. Zarif Shaukat Ansari, Ms. Anagha Mahadeo Jagtap , Ms. Shilpa Suresh Raut, "Data Leakage Detection and E-mail Filtering", International Journal of Innovative Research in Computer and Communication Engineering, Volume-1,Issue-3,May 2013.

[7]. Prof. Shushilkumar N. Holambe, Dr. Ulhas B. Shinde, Archana U. Bhosale, "Data Leakage

using cloud Computing", International Journal of Scientific and Engineering and Research, Volume 6, Issue4, April 2015.

[8]. V. Shobana, M. Shanmugasundaram, "Data Leakage Detection using Cloud Computing", International Journal of Emerging Technology and Advanced Engineering ICISC, Volume 3,Issue 1,January 2013.

[9]. Chandani Bhatt, Prof. Richa Sharma , "Data Leakage Detection" ,International Journal of Computer Science and InformationTechnologies(IJCSIT),Vol.5(2),2014

[10]. D. Kumuthavijay, J. Nandhini, V. Jayaprakasan, "Implementation of Efficient Audit Service Outsourcing for Data Integrity by Interfacing the Mobile device in Cloud", International journal of Computer Application ,Volume 67-No.20,April 2013.

[11]. Chandu Vaidya and Prashant Khobragade , "Data Security in Cloud Computing", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN:2321-8169,Volume 3,Issue 5,pp-167-170.

[12]. Chandu Vaidya, Prashant Khobragade and Ashish Golghate, "Data Leakage Detection and Security in Cloud Computing", GRD Journals-Global Research Development Journal for Engineering,Volume 1,Issue 12,November 2016.