Global Journal of Computer Science and Technology: B
Cloud and Distributed
Volume 14 Issue 3 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Data Leakage Detection using Cloud Computing

By Sandilya Pemmaraju, V. Sushma & Dr. K. V. Daya.Sagar
K L University, India

Abstract- Today the present world mostly depends on exchange of information i.e. transfer of data from one person to another person which is also known as distributary system. The data is sent from the distributor to the user are confidential so the data is distributed only between the distributor and the trusted third parties. The data sent by the distributor must be secured, confidential and must not be reproduced as the data shared with the trusted third parties are confidential and highly important. In some occasions the data distributed by the distributor are copied by different agents who cause a huge damage to the institute and this process of losing the data is known as data leakage. The data leakage must be detected in the early stage in order to protect the data form being open source. This project deals with protecting the data from being out sourcing by giving a special inscription to the sensitive data so that it cannot be reproduced.

GJCST-B Classification: C.2.4

Strictly as per the compliance and regulations of:

Data Leakage Detection using Cloud Computing

Sandilya Pemmaraju [α], V. Sushma [σ] & Dr. K. V. Daya.Sagar [ρ]

Abstract- Today the present world mostly depends on exchange of information i.e. transfer of data from one person to another person which is also known as distributary system. The data is sent from the distributor to the user are confidential so the data is distributed only between the distributor and the trusted third parties. The data sent by the distributor must be secured, confidential and must not be reproduced as the data shared with the trusted third parties are confidential and highly important. In some occasions the data distributed by the distributor are copied by different agents who cause a huge damage to the institute and this process of losing the data is known as data leakage. The data leakage must be detected in the early stage in order to protect the data form being open source. This project deals with protecting the data from being out sourcing by giving a special inscription to the sensitive data so that it cannot be reproduced.

I. Introduction

Every company focus on security issues of securing the data from different third parties form being out sourced. Every company follows a different strategy which does not match with any other company. The employees are also trained in order to maintain the secrecy of the data and maintain the basic structure of the company. The security must be beyond the employees' knowledge so that the employee has no idea of cracking it by covering logical and physical security.

Information security is frequently subjected to metaphors. The information security must be targeted at global level by not letting the user know the problematic issues faced by the security department and also the logical security i.e. the sensitive data, applications and also the operating system used in a particular institute. The security must also be extended to telecommunication department of the institute so that they have a network security also.

There is no particular period of data leakage it may happen at any time. Data leakage only depends on the importance of the information distributed by the distributor. The information distributed is considered as sensitive data when it consists of information about the client, budget, code and any design specification. If the leakage occurs it leaves the institute in unprotected state. This data leakage puts the institute in ambiguity which results in the downgrade of the business and ultimately failure of the company.

The agents who get their hands on the sensitive data are also known as cyber criminals. Data leakage is done for their own profits which results in loss of the company. To overcome this problem we tried a new idea of adding fake objects to the distributary data to find the agent who are misusing the data and take certain actions. The information consists of one constrains and one fake object to meet the requirements of the security.

II. Water Marking the Information

Water marking is a type of security technique which deals with the idea of embedding a particular code or encryption on the information that is to be distributed. The information can be image or a video or any official file. This encryption helps the company to claim the ownership on any particular data.

Water marking is a technique where a bit pattern is added to the data at a particular position on the tuples and subset of the data. The tuple and subset and their attributes are algorithmically coded in such a way that they are controlled by a key which can be accessed only by the owner

We don't need to have the access to the original data or the pattern of watermark to detect the watermark. The watermark can be detected from a small subset of data which consists of a small portion of water mark. The watermarking is installed into the file by using watermarking software which feds the information with small errors. The combination of those small error forms into a watermark which does not have any significant meaning and they cannot be destroyed by the external source.

While coming to the digital data i.e. images, videos and documents can easily be leaked and can go vital on the web and this can be efficiently tackled by using watermarking as a proof of ownership by using a signature.

III. Need for Data Allocation

The process of allowing one company to access the data of another company is known as information system and it is very essential for companies

Author α: IV B.Tech ECM, K L University.
e-mail: Sandilya.pemmaraju@gmail.com
Author σ: IV B.Tech ECM, K L University.
e-mail: jwsboardfe @gmail.com
Author ρ: Associate Professor, Department of Computer Science &Engineering, K L University. e-mail: sagarTadepalli@ gmail.com.

and must be preserved with high security. Security issues generally consists of monitoring the software and materials provided by the institute are only used in the institute and also provide data privacy and maintain the relation between gathering and disclosing of data with all legal and political issues.

The data is open only when the organization sends secure information to untrusted third party unintentionally considering it as a trusty third party. The distributor later discovers that some set of objects are sent to a new location and the data is being leaked and now the goal is to locate the agent and oppose the agent from accessing the data. We must not only stop the access to the data but also check whether the agent is a leaker or not.

The distributor sends the data to the agent using data allocation strategies to increase the possibility of finding the agent by adding fake objects to the information distributed. If any person receiving the data leaks the data then the distributor will find the agent by the help of numbers of fake objects released out and the distributor waits until he gets enough evidence and finally conform the agent and closes the business with him or takes any legal action on the agent.

IV. Cloud Computing Introduction

The interconnection of large number of system virtual is known as cloud. The interconnected systems may be private or public and the data stored in it can also be differentiated into private and public access. For example one drive powered by Microsoft is one of the examples of cloud computing. In one drive every user is given with certain user id and password and can access the data in the cloud from anywhere by just connecting to internet and using their unique user id and password.

The cloud computing gives access to a wide access throughout the world. Any person with an authorised id can access the information of the organization from any part of the world from any computer through internet. The infrastructure of cloud does not contain any physical data all the data stored by the user is stored virtually by using cloud server which are maintained by using HTML and XML code.

According to Google there are six properties of cloud computing, they are

- User-centric: When a person access the cloud whatever information stored in the cloud can be accesses by him and the information stored by him can be accessed by others if the user gives permission
- Task-centric: It shares the documents files and folders stored in the cloud only to the authorised persons i.e. to the person with whom the admin shared his rites.
- Powerful: It can connect hundreds of system at a single time and can give access to a single file to many systems which cannot be done by any sharing device.
- Accessible: User can easily access the data stored in the cloud by simply accessing his user id from any system which makes it more flexible.
- Intelligent: With all the different data stored in the cloud it easily does the mining activity and displays the result when the user searches.
- Programmable: The task handled by the cloud must be automated

V. Related Work

Reference Paper 1: Rights Protection is provided for Relational Data

This paper deals with the idea of creating bit patterns on the file at certain location and all the bit patterns combined and make a watermark. The bits entered are set of numbers which provide right protection to the data that is present in the data base.

This paper also deals with the development of watermark detection application which reads the algorithms of the bit pattern by locating the markings and retrieves the original data at the client side.

The main failure in this model is that it only deals with the numerical encoding of the data and does not detect the nonnumeric data in the database

Reference paper 2: Watermarking Technique for Multimedia Data

This paper was developed on the technique of watermarking the data using multi-media watermarking technology to prevent the digital content going vital on net by disabling the copy facility.

Encryption of the data has its own limitation from protecting the information. If the rights are decrypted then the data cannot be protected from illegally replicating the digital content.

But this encryption problem is overcome by sung digital watermark which is embedded on the host data and cannot be removed and it includes the copyrights, data protection and monitoring and tracking.

Reference Paper 3: Achieving K-Anonymity Privacy Protection

This paper deals with generalization and suppression techniques to protect the data from leakage using K-anonymity privacy protection. Where every part of the data is divided into k different subsets and every subset is linked with specific set of details and the final data is obtained at the external source.

This technique is a failure as it lacks in clear description on how the data is being secured and what happens to the data if they are not systematically liked to one another

VI. Conclusion

The data leakage detection in information system is obtained by following basic strategies like watermarking on different information. This data leakage issue can be handled in multiple ways which must be studied later.

When the information is watermarked it secures the data from being open source and helps to find out the cybercriminal by using the fake objects placed at different positions of information that is to be sent.

References Références Referencias

1. R. Sion, M. Atallah, and S. Prabhakar, •\Rights Protection forRelational Data,. Proc. ACM SIGMOD, pp. 98-109, 2003.
2. Hartung and kutter,.Watermarking technique for multimedia data.2003.
3. L. Sweeney, —Achieving K-Anonymity Privacy Protection Using Generalization And Suppression, http://en.scientificcommons.org/43196131, 2002.

This page is intentionally left blank