INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

DATA LEAKAGE DETECTION IN CLOUD COMPUTING

GAURAV GULHANE, PRASHANT KHOBRAGADE, ASHISH GOLGHATE, CHANDU VAIDYA, ARPIT CHAUDHARI

Department of Computer Science & Engg., Dr. R. G. I. T. & R., Amravati.

Accepted Date: 15/03/2016; Published Date: 01/05/2016

Abstract: Cloud computing is a network based computing with the intention to store data over the internet. Cloud computing is a place that can store numerous scale data and has various computing techniques. It facilitates computing with faultless access to virtually limitless resources. Such computing techniques have tremendous advantages such as cost effectiveness, increased storage, shared resources, easy backup and restoration etc. Although there would be some challenges at some point of data sharing in the cloud computing namely data hijacking, data leakages etc. The main objective of this context is to spotlight the issue of data leakage and storage services, probable to be data leakages in the cloud storage services. This paper initiates a scheme, impeding Data Leakage technique in the cloud computing which is involved in hampering the leakage of data due to intrusions in the transmission pathway and unauthorized permission granted to the different services to access the cloud. It leads to halt the entire data leakage with that meticulous path. The objective of this paper is to identify the data leakage in cloud computing and the different techniques used to overcome this data leakage in storage over the cloud.

Keywords: Cloud Computing, Data leakage, Data leakage hindering, Data leakage detection

Corresponding Author: MR. GAURAV GULHANE

Access Online On:

www.ijpret.com

How to Cite This Article:

Gaurav Gulhane, IJPRET, 2016; Volume 4(9): 1610-1616

PAPER-QR CODE

INTRODUCTION

Data leakage is the unauthorized transfer of confidential information from a computer or datacenter to the outside world. Data leakage can be done by simply remembering what was seen, by physical tampered of tape, disks and reports or by slight means such as data hiding. Data leakage is a problem for anyone that uses a computer and data store on some sever side. Data leakage happens when data has been tampering either intentionally or unintentionally. The data leakage has a biggest problem in organization, where the organizations are in responsibility try to overcome this problem. Data Leakage is an incident when the confidentiality of information has been compromised. It refers to an unauthorized transmission of data from one party to other or within an organization to an outside destination. The data that is leaked out can either be private in nature and are deemed confidential whereas Data Loss is loss of data due to deletion, system crash, in-depended storage, services provided to access it etc. Totally both the term can be referred as data breach, has been one of the biggest fears that organization face today [1].

Dropbox allows users to create a special folder on their computers, which Dropbox then synchronizes so that it appears to be the same folder (with the same contents) regardless of which computer is used to view it. Files placed in this folder are also accessible via the Dropbox website and mobile apps. Dropbox uses a freemium business model, where users are offered a free account with a set storage size and paid subscriptions for accounts with more capacity.

When a few hundred Dropbox users began receiving spam emails about online casinos and gambling sites two weeks ago, it seemed like something was up. And indeed there was. The online file storage service confirmed today that hackers accessed usernames and passwords from third party sites and then used them to get into Dropbox users accounts.

The investigator found that usernames and passwords recently stolen from other websites were used to sign in to a small number of Dropbox accounts[9]. They contacted these users and helped them protect their accounts, the company wrote in a blog post today. "A stolen password was also used to access an employee Dropbox account containing a project document with user email addresses. We believe this improper access is what led to the data leakage. These are the main reason behind the data leakage detection and dependable storage services. Data leakage detection and dependable storage services not only keep our data safe but also helps us to be tension free [4].

Data leakage is not a problem only associated with the cloud. Hackers, spyware and inadvertent data breaches can also lead to data leakage from a company server. Data loss is self-explanatory

and it can happen if someone losses a briefcase. But, on a business scale, data loss can be disastrous to the business and its clients. Most important of them all is that, there should be a good degree of encryption provided by the vendor to the user that only the user should be able to access the data and not the malicious user [2].

Figure 1: Cloud Computing

Mobile devices present yet another challenge for data leakage. USB keys, Bluetooth devices or removable CD drives, for example, can all circumvent network controls without a system administrator's knowledge. As hardware storage devices, they outdo the sophisticated Internet and Web-monitoring tools just described [4].

One such tool, Safend Protector V3.0, can be installed as a client on all the desktops and laptops in your enterprise. It can be centrally managed via a Web-based interface and, like the Web monitoring tools, can be tuned to check for certain types of data being moved through USB, Firewire or wireless ports.[8] The tool is tamper-proof, invisible to users, and silent until something is connected to an external port. Additionally, Safend Protector V3.0 can be tuned to completely block access to any removable device, restrict certain devices based on capacity, or allow read-only access and policies can integrated into the Group Policy Objects (GPO) of Active Directory to provide access to devices for selected users.

The rest of the paper organized as follows: Section 2 literature surveys related to put this paper in the right context. Section 3 describes the proposed method for data leakage detection. In section 4 describe methodology used for identification of data leakage. Section 5draws conclusions and future work.

II. LITERATURE SURVEY

Rights protection for relational data is of ever increasing interest, especially considering areas where sensitive, valuable content is to be outsourced. It handles data security through

watermarking in the framework of numeric relational data and instead of primary key it uses the most significant bits of the normalized data set. Mainly, it divides the data set into partitions using markers and then varies the partition statistics to hide watermark bits [1].

It proposes a watermark embedding algorithm such that it consists of Sorting, Partitioning used for marker location and bit embedding watermark bits are embedded in the number set so as to provide a right protection to the data that are stored into it the relational database.

The major drawback is that it should not deal on the area of data security through watermarking in the framework of nonnumeric encoding domains in this relational database [1].

Hartung and Kuttur [2] focus on the Multimedia watermarking technology that has evolved very quickly during the last few years. Encryption technologies can be used to prevent unauthorized access to digital content, [7]it is clear that encryption has its limitations in protecting intellectual property rights once content is decrypted, and there's nothing to prevent an authorized user form illegally replicating digital content. It provides the requirements and all the related applications for watermarking is reviewed. The application includes copyright protection, data monitoring, and data tracking. Robustness and security aspects are also discussed in specific data source.

Latanya Sweeney [3] deals about generalization and suppression techniques to safeguard the data from the data distributors using k-anonymity privacy protection. The data in the system is analyzed for generalization like replacing or recoding a value with a less specific but semantically consistent values and suppression involves not releasing a value at all. It mainly focused towards suppression technique which is nothing but it should not provide the data to the user [4].The major drawback in this system is that, there is no clear explanation on, how the data is going to be secured in suppression technique. The next issue is, by considering the data when it is not semantically linked then the suppression technique should not be effective.

III. PROPOSED SYSTEM

The aim of the data leakage detection and dependable storage service is to secure cloud computing system which can detect data loss, data leakage or any events questioning data integrity with a user friendly interface.

Our goal is to detect when the distributor's sensitive data has been leaked by agents, and if possible to identify the agent that leaked the data.

A data distributor has given sensitive data to a set of supposedly trusted agents (third parties). Some of the data is leaked and found in an unauthorized place (e.g., on the web or somebody's laptop). The distributor must assess the likelihood that the leaked data came from one or more agents, as opposed to having been independently gathered by other means.

We propose data allocation strategies (across the agents) that improve the probability of identifying leakages. These methods do not rely on alterations of the released data (e.g., watermarks). In some cases we can also inject "realistic but fake" data records to further improve our chances of detecting leakage and identifying the guilty party.

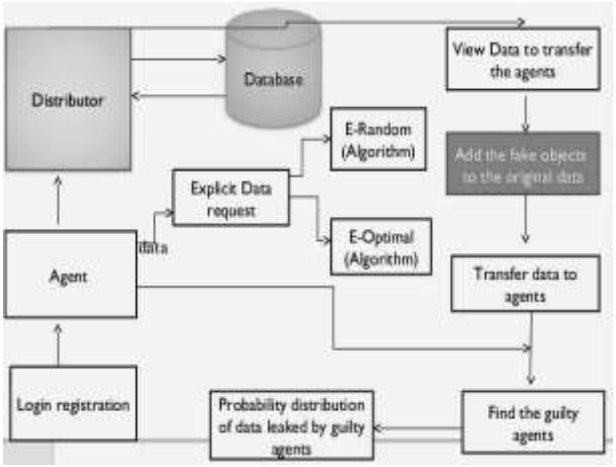

Figure2: Block diagram of Proposed approach

Agent: Distributer Relation

• 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 distributaries 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.

Problems with supposedly trusted agents

• Since users may not retain a local copy of outsourced data, there exist various incentives for cloud service providers to behave unfaithfully towards the cloud users regarding the status of their outsourced data.

• They may reduce the data, without the user's permission or they may leak the data which will be a question to data integrity.

IV: METHODOLOGY

Data leakage detection with dependable storage in cloud with sensitive data identification used to identify where the data is access without the prior permission of user. The following methods are used for this:

1) Storage verification schemes:

• The Hash table Storage of the outsourced data can be maintained at the user's end.

• The name and the size of the file can be taken as hash keys.

• Using this table, verification of the outsourced data can be done by generating the key and matching with the user's hash table.

2) Watermarking:

• In watermarking a unique code is embedded in each distributed copy.

• If that copy is later discovered in the hands of an unauthorized party, the leaker can be identified.

• Watermarking techniques provide unique identification to the data

• Details of value of each of the pixel of data or image, if any intended [6].

CONCLUSION

This paper proposed a scheme, impeding Data Leakage technique in the cloud computing which is involved in hampering the leakage of data due to intrusions in the transmission pathway and unauthorized permission granted to the different services to access the cloud. It leads to halt the entire data leakage with that meticulous path.

REFERENCE

1. R. Sion, M. Atallah, and S. Prabhakar, Rights Protection forRelational Data,‖ Proc. ACM SIGMOD, pp. 98-109, 2003.
2. R.Agrawal and J. Kiernan, —Watermarking relational databases‖. In VLDB ’02: Proceedings of the 28th international conference on Very Large Data Bases, pages 155–166. VLDB Endowment, 2002.
3. Hartung and kutter,‖Watermarking technique for multimedia data‖2003.
4. Chandu Vaidya, Prashant Khobragade, “Data Security in Cloud Computing”, International Journal on Recent and Innovation Trends in Computing and Communication, Vol. 3, Issue: 5, May 2015.
5. Y. Cui and J. Widom, "Lineage Tracing for General Data Warehouse Transformations," The VLDB J., vol. 12, pp. 41-58, 2003. [CrossRef]
6. S. Czerwinski, R. Fromm and T. Hodes, "Digital Music Distribution and Audio Watermarking,"http://www.scientificcommons. org/43025658, 2007.
7. F. Guo, J. Wang, Z. Zhang, X. Ye and D. Li, "An Improved Algorithm to Watermark Numeric Relational Data," Information Security Applications, pp. 138-149, Springer, 2006. [CrossRef]
8. F. Hartung and B. Girod, "Watermarking of Uncompressed and Compressed Video," Signal Processing, vol. 66, no. 3, pp. 283-301, 1998.
9. P. Papadimitriou and H. Garcia-Molina, “Data leakage detection,” IEEE Transactions on Knowledge and Data Engineering, pages 51-63, volume 23, 2011.