WWJMRD 2018; 4(3): 241-244
www.wwjmrd.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
UGC Approved Journal
Impact Factor MJIF: 4.25
E-ISSN: 2454-6615

Gagandeep Kaur
Research scholar, University College of Computer Applications, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

Dr.Sandeep Kautish
Professor in Computer Science, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

Correspondence:
Gagandeep Kaur
Research scholar, University College of Computer Applications, Guru Kashi University, Talwandi Sabo, Bathinda, Punjab, India

A Review on Data Leakage Detection & Prevention Using Cloud Computing

Gagandeep Kaur, Dr.Sandeep Kautish

Abstract
In today's business world, many organizations use Information Systems to manage their sensitive and business critical information. The need to protect such a key component of the organization cannot be over emphasized. Data Loss/Leakage Prevention has been found to be one of the effective ways of preventing Data Loss. Data Loss Prevention is found to be the data leakage/loss control mechanism that fits naturally with the organizational structure of businesses. It not only helps the organization protect structured data but it also helps protection and leakage prevention of unstructured data.
DLP is considered as preventive control which when applied helps organization prevents data leakage of sensitive information (Personal identifiable information, financial information, trade secrets, merger and acquisitions etc.). The DLP solution is not only for the big organizations and for particular industry sector like banking and finance but it is a need for small organizations and other fields of business (Health care, aviation, consulting etc.) due to various Laws and Regulatory requirement by different countries. In this paper we study the various mechanisms for data leakage detection and prevention**.**

Keywords: Data Leakage Detection & Prevention; DLP; Cloud Computing; Security in Cloud Computing

I. Introduction

Data loss, which means a loss of data that occur on any device that stores data. It is a problem for anyone that uses a computer. Data loss happens when data may be physically or logically removed from the organization either intentionally or unintentionally. The data loss has become a biggest problem in organization today where the organizations are in responsibility 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 within an organization to an external 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 etc. Totally both the term can be referred as data breach, has been one of the biggest fears that organization face today.

Data Loss/Leakage Prevention (DLP) is a computer security term which is used to identify, monitor, and protect data in use, data in motion, and data at rest[1]. DLP is sued to identify sensitive content by using deep content analysis to per inside files and with the use if network communications. DLP is mainly designed to protect information assets in minimal interference in business processes. It also enforces protective controls to prevent unwanted incidents. DLP can also be used to reduce risk, and to improve data management practices and even lower compliance cost.

Systems are designed to detect and prevent unauthorized use and transmission of confidential information. Vendors refer to the term as Data Leak Prevention, Information Leak Detection and Prevention (ILDP), Information Leak Prevention (ILP), Content Monitoring and Filtering (CMF), Information Protection and Control (IPC)

or Extrusion Prevention System by analogy to Intrusion-prevention system[1].

A. Advantages of Cloud DLP

In a cloud environment, a virtual machine can be used to run a security engine in order to manage all the other virtual machines on a designated set of virtual servers, based on virtual machine manager technology to host virtual machines. The virtual machines can then run client software with a DLP engine that will scan, recognize and block communication of sensitive information. The VMM can get these together and merge into a single virtual machine, making DLP engine able to monitor and manage all the virtual machines that run a client, and also to see data at rest. This makes the scope for compliance requirements like PCI DSS; PII etc. for sensitive data. DLP runs as a service, it can be enabled / disabled for virtual machines running in the cloud data center.

A cloud environment is dynamic, so as a DLP service, as it can be extensible and automated. A DLP solution can be planned using APIs to automate controls, like making a rule that automatically shift a virtual machine with sensitive data behind a firewall or budge it into a lockdown.

The flexibility and control in the cloud computing makes control of virtual machines more viable than in the physical setup. A rule can require a VM found with credit card data, should have its network connectivity isolated at the application level (restrict certain protocols) to block data leaks, and shoot an alert (email) to administrators. Assessment of a full virtual data center; cloud

DLP can find systems with sensitive data and move them from a cluster of insecure systems to

One assigned to business-critical applications with sensitive data.

B. Limitations of Cloud DLP

If the cloud platform is public it may support a single network interface per instance, which will result in a need of virtual DLP version that can monitor and forward or block traffic with restriction. There is a lot of significance in using DLP to monitor data migrating to the cloud and for content discovery on cloud-based storage, but deploying DLP in a public cloud may not be significant. It makes sense in private cloud, depending on what it is used for. Security of any cloud deployment in line with DLP is probably an application infrastructure, which rely more on application security and encryption.

DLP is an excellent tool to enhance data security in the cloud. It can be used to track data migrating to the cloud, discover sensitive information stored on cloud, and to protect services running on the cloud, given the fact it is tuned accordingly

II. Literature Survey

Jayasimha S.R [1], Cloud computing is a collection of large group of remote-servers which allow centralized data storage, access to resource. Resources are pooled and pay as you need the resource. The services provided are data storage, networking and application software's. This paper focus on the security issues in the Cloud storage. Different models in cloud are private cloud, public cloud and Hybrid cloud. Three tier-setup can be done.i.e. i) Users of virtual services ii) tenants who provide services iii) providers who provide the infrastructure. In gaining the trust in cloud computing one should ensure the proper security such that data will not be leaked in between the layers. Data can be leaked from one to many. So, we have to prevent the data from identified leakage steps. Data leakage is found in unauthorized places e.g.1) E-mail leak scenario while data transfer using SMTP server [1]. 2) From laptops Data Leak Prevention is a technique to detect data leakage and monitoring and blocking data which are in use, in motion and at rest. DLP technique consistently monitors the data flow which helps in detecting and controlling the data leakages and MYDLP is open source software is used to prevent the data leakage from the data base.

Sandip A. Kale [2], this paper contains the results of implementation of Data Leakage Detection Model. Currently watermarking technology is being used for the data protection. But this technology doesn't provide the complete security against date leakage. This paper includes the difference between the watermarking & data leakage detection model's technology. This paper leads for the new technique of research for secured data transmission & detection, if it gets leaked.

S. Pemmaraju [3], 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.

Te-Shun Chou [4], Clouds provide a powerful computing platform that enables individuals and organizations to perform variety levels of tasks such as: use of online storage space, adoption of business applications, development of customized computer software, and creation of a "realistic" network environment. In previous years, the number of people using cloud services has dramatically increased and lots of data has been stored in cloud computing environments. In the meantime, data breaches to cloud services are also increasing every year due to hackers who are always trying to exploit the security vulnerabilities of the architecture of cloud. In this paper, three cloud service models were compared; cloud security risks and threats were investigated based on the nature of the cloud service models. Real world cloud attacks were included to demonstrate the techniques that hackers used against cloud computing

systems. In addition, countermeasures to cloud security breaches are presented.

III. Data Leakage Detection Modules

I. Data Allocation Module

The main focus of our project is the data allocation problem as how can the distributor "intelligently" give data to agents in order to improve the chances of detecting a guilty agent, Admin can send the files to the authenticated user, users can edit their account details etc. Agent views the secret key details through mail. In order to increase the chances of detecting agents that leak data.

II. Fake Object Module

The distributor creates and adds fake objects to the data that he distributes to agents. Fake objects are objects generated by the distributor in order to increase the chances of detecting agents that leak data. The distributor may be able to add fake objects to the distributed data in order to improve his effectiveness in detecting guilty agents. Our use of fake objects is inspired by the use of "trace" records in mailing lists. In case we give the wrong secret key to download the file, the duplicate file is opened, and that fake details also send the mail. Ex: The fake object details will display.

III. Optimization Module

The Optimization Module is the distributor's data allocation to agents has one constraint and one objective. The agent's constraint is to satisfy distributor's requests, by providing them with the number of objects they request or with all available objects that satisfy their conditions. His objective is

To be able to detect an agent who leaks any portion of his data. User can able to lock and unlock the files for secure

IV. Data Distributor Module

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 Admin can able to view the which file is leaking and fake user's details also.

V. Agent Guilt Module

To compute this, we need an estimate for the probability that values in S can be "guessed" by the target. For instance, say some of the objects in T are emails of individuals. We can conduct an experiment and ask a person with approximately the expertise and resources of the target to find the email of say 100individuals. If this person can find say 90 emails, then we can reasonably guess that the probability of finding one email is 0.9. On the other hand, if the objects in questionnaire bank account numbers, the person may only discover say 20, leading to an estimate of 0.2. We call this estimate pt, the probability that object t can be guessed by the target. To simplify the formulas that we present in the rest of the paper, we assume that all T objects have the same pt, which we call p. Our equations can be easily generalized to diverse pt though they become cumbersome to display. Next, we make two assumptions regarding the relationship among the various leakage events. The first assumption simply states that an agent's decision to leak an object is not related to other objects.

VI. Existing Techniques for data leakage detection and prevention

A. Watermarking Technique

Traditionally, leakage detection is handled by watermarking, e.g., 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. Watermarks can be very useful in some cases, but again, involve some modification of the original data. Furthermore, watermarks can sometimes be destroyed if the data recipient is malicious. E.g. A hospital may give patient records to research who will devise new treatments. Similarly, a company may have partnerships with other companies that require sharing customer data. Another enterprise may outsource its data processing, so data must be given to various other companies. We call the owner of the data the distributor and the supposedly trusted third parties the agents.

B. Data Allocation Strategy

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. We also present algorithm for distributing object to agent. 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. Perturbation is a very useful technique where the data is modified and made 'less sensitive' before being handed to agents. We develop unobtrusive techniques for detecting leakage of a set of objects or records. In this section we develop a model for assessing the 'guilt' of agents. We also present algorithms for distributing objects to agents, in a way that improves our chances of identifying a leaker. Finally, we also consider the option of adding 'fake' objects to the distributed set. Such objects do not correspond to real entities but appear realistic to the agents. In a sense, the fake objects acts as a type of watermark for the entire set, without modifying any individual members. If it turns out an agent was given one or more fake objects that were leaked, then the distributor can be more confident that agent was guilty. Today the advancement in technology made the watermarking system a simple technique of data authorization. There are various software which can remove the watermark from the data and makes the data as original.

C. Data leakage detection using Cloud Computing

In the virtual and widely distributed network, the process of handover sensitive data from the distributor

to the trusted third parties always occurs regularly in this modern world. It needs to safeguard the security and durability of service based on the demand of users. The idea of modifying the data itself to detect the leakage is not a new approach. Generally, the sensitive data are leaked by the agents, and the specific agent is responsible for the leaked data should always be detected at an early stage. Thus, the detection of data from the distributor to agents is mandatory. This project presents a data leakage detection system using various allocation strategies and which assess the likelihood that the leaked data came from one or more agents. For secure transactions, allowing only authorized

users to access sensitive data through access control policies shall prevent data leakage by sharing information only with trusted parties and also the data should be detected from leaking by means of adding fake record`s in the data set and which improves probability of identifying leakages in the system. Then, finally it is decided to implement this mechanism on a cloud server Key to the definition of cloud computing is the ―cloud itself. For our purposes, the cloud is a large group of interconnected computers. These computers can be personal computers or network servers; they can be public or private. For example, Google hosts a cloud that consists of both smallish PCs and larger servers. Google's cloud is a private on (that is, Google owns it) that is publicly accessible (by Google's users).This cloud of computers extends beyond a single company or enterprise. The applications and data served by the cloud are available to broad group of users, cross enterprise and cross-platform. Access is via the Internet. Any authorized user can access these docs and apps from any computer over any Internet connection. And, to the user, the technology and infrastructure behind the cloud is invisible. It isn't apparent (and, in most cases doesn't matter) whether cloud services are based on HTTP, HTML, XML, Java script, or other specific technologies. From Google's perspective, there are six key properties of cloud computing.

Conclusion

From the above discussion it is concluded that the data leakage detection industry is very heterogeneous as it evolved out of ripe product lines of leading IT security vendors. A broad arsenal of enabling technologies such as firewalls, encryption, access control, identity management, machine learning content/context-based detectors and others have already been incorporated to offer protection against various facets of the data leakage threat. The competitive benefits of developing a "one-stop-shop", silver bullet data leakage detection suite is

mainly in facilitating effective orchestration of the aforementioned enabling technologies to provide the highest degree of protection by ensuring an optimal fit of specific data leakage detection technologies with the "threat landscape" they operate in. This landscape is characterized by types of leakage channels, data states, users, and IT platforms.

References

1. Jayasimha S.R, Narasimha Prasad D,Hamsa K and Hamsa K,"Prevention of Data from the Data Leakage in Cloud Computing", Prevention of Data from the Data Leakage in Cloud Computing.
2. Sandip A. Kale, Prof. S.V.Kulkarni,"Data Leakage Detection", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 9, November 2012.
3. Sandilya Pemmaraju, V. Sushma & Dr. K. V. Daya.Sagar,"Data Leakage Detection using Cloud Computing", Global Journal of Computer Science and Technology: B Cloud and Distributed Volume 14 Issue 3 Version 1.0 Year 2014.
4. Te-Shun Chou,"SECURITY THREATS ON CLOUD COMPUTING VULNERABILITIES", International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 3, June 2013.
5. Rudrani Sinha, Prof. Chaitali Choudhary,"Information Leak Detection System using Fingerprint of data", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 2 Issue: 12.
6. K. Manoj Kumar, G. Shubhang, G. Rajesh Chandra,"DATA LEAKAGE DETECTION SYSTEM FOR CLOUD-BASED STORAGE SYSTEMS", (IJAER) 2014, Vol. No. 8, Issue No. V, November.
7. Bijayalaxmi Purohit, Pawan Prakash Singh,"Data leakage analysis on cloud computing", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622.
8. Bhatt, Prof.Richa Sharma, "Data Leakage Detection", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (2), 2014, 2556-2558.
9. Suresh Kumar, K. Iyakutti,"A novel Technique: Data Leakage Hindering in Cloud computing using Swarm Intelligence", Int.J.Computer Technology & Applications, Vol 5 (6), 1886-1891.
10. Shobana, M. Shanmugasundaram, "DATA LEAKAGE DETECTION USING CLOUD COMPUTING",International Conference on Information Systems and Computing (ICISC-2013), INDIA.
11. Alex Ofori Karikari, Joseph Kobina Panford, James Ben Hayfron-Acquah, Frimpong Twum, "Detecting Data Leakage in Cloud Computing Environment", International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-3, June2015G. Eason, B. Noble, and I.N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955