Volume 6, Issue 2, February 2016 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

# Effectiveness of Data Loss Prevention in Cloud Computing

**Dipali Pattanayak**[*]
Assistant Professor, Department of IT
BVRIT College of Engineering for Women,
Hyderabad, India

**Amarinder Kaur**
Assistant Professor, Department of CSE
BVRIT College of Engineering for Women,
Hyderabad, India

***Abstract— In this era, cloud computing is the biggest buzz in ITworld. Cloud computing is an Internet-based computing, where shared resources, software and information, are provided to computers and devices on-demand. Due to its perceived open nature, it raises strong security, privacy and trust concerns. Unfortunately, the adoption of cloud computing speeded before appropriate technologies appeared to tackle the accompanying challenges of trust. While accessing shared resources in cloud data must address security and privacy, especially when it comes to managing sensitive data. In this paper we have discussed data loss as a big threat in cloud computing .We have also focused different Data Loss Prevention approaches to resolve this problem.***

***Keywords— DLP, SaaS, Dataloss***

## I. INTRODUCTION

Technologies such as cluster, grid, and now, cloud computing, have all aimed at allowing access to large amounts of computing power in a fully virtualized manner, by aggregating resources and offering a single system view. In addition, an important aim of these technologies has been delivering computing as a utility. Utility computing describes a business model for on-demand delivery of computing power; consumers pay providers based on usage (“pay as-you-go”), similar to the way in which we currently obtain services from traditional public utility services such as water, electricity, gas, and telephony.

Cloud computing has been coined as an umbrella term to describe a category of sophisticated on-demand computing services initially offered by commercial providers, such as Amazon, Google, and Microsoft. It denotes a model on which a computing infrastructure is viewed as a “cloud,” from which businesses and individuals access applications from anywhere in the world on demand .The main principle behind this model is offering computing, storage and software “as a service.”[1]

Many of the features that make cloud computing attractive, have not just challenged the existing security system, but have also revealed new security issues. Due to its perceived open nature, Cloud Computing raises strong security, privacy and trust concerns. Unfortunately, the adoption of cloud computing came before the appropriate technologies appeared to tackle the accompanying challenges of trust. This gap between adoption and innovation is so wide that cloud computing consumers don’t fully trust this new way of computing. To close this gap, we need to understand the trust issues associated with cloud computing from both a technology and business perspective.

However, customers are also very concerned about the risks of Cloud Computing if not properly secured, and the loss of direct control over systems for which they are nonetheless accountable .There are different security issues identified by the consumers like : Insecure Interfaces and APIs, Malicious Insiders, Data loss or leakage, Shared Technology Issues and many more.

In this research paper our main focus is on data loss in cloud computing. Here we have discussed about traditional DLP,how cloud DLP is different than traditional one.We have mentioned three different approaches of cloud DLP to prevent dataloss in cloud computing.

## II. WHAT IS DATA LOSS?

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.The data loss issue is being exposed from confidential information about a customer to dozens of company’s product files and documents being sent to a competitor. This can be caused in many ways either accidental or deliberate, or even with insiders in realizing sensitive data about customer’s personal information, intellectual property, or other confidential information in violation of company policies and regulatory requirements.[3] In organization, today’s employees with available access to electronically expose sensitive data, the scope of sensitive data loss problem is greater than outsider’s threat protection. In order to cover all the loss bearings, an organization has the potential to encounter:

A. *Data in motion* – Any data that is moving through the network to the outside via the

B. *Internet Data at rest* – Data that resides in files systems, databases and other storage methods

C. *Data at the endpoint* – Data at the endpoints of the network (e.g. data on devices such as USB, external drives, laptops, mobile devices, etc)[8,9].

## III. DATA LOSS IN CLOUD COMPUTING

In cloud computing, for both consumers and businesses, the prospect of permanently losing one's data is terrifying. For example in 2012, attackers broke into Mat's Apple, Gmail and Twitter accounts. They then used that access to erase all of his personal data in those accounts. Permanently loosing data hosted on a cloud can result from several reasons like:

A. Attack by Malicious hackers.
B. Any accidental deletion by the cloud service provider.
C. A physical catastrophe such as a fire or earthquake, could lead to the permanent loss of customers' data unless the provider takes adequate measures to backup data.

## IV. TRADITIONAL DLP

The term DLP stands for Data Loss/Leakage Prevention, which was introduced in 2006. Gradually DLP has been improved and became a strong mechanism to influence the security industry worldwide. DLP is used to bear Information loss Prevention/Protection, Information Leak prevention/Protection, and Extrusion Prevention. Later this DLP technology gained some popularity in the early year 2007.[10]

### *A. DLP key features*

DLP allows enterprise to:

1. *Monitoring* - DLP identifies a wide range of sensitive enterprise content, from information in confidential documents, to customer and privacy related information, to content specified by customers, or provided out-of-the-box.
2. *Enforcing* - DLP uses information gathered from monitoring to enforce enterprise data privacy policies and to meet designated compliance requirements.
3. *Analysis* - DLP recognizes over 900 different file types. Analysis of the data is based on the actual content of the file and not the extension that is used with the file.[7]

## V. SHORTCOMINGS OF TRADITIONAL DLP IN CLOUD

A. *Lack of basic visibility*: They can only monitor traffic on enterprise controlled assets (e.g., networks/endpoints). However, traffic to and/from a SaaS application might not go over an enterprise network at all.

B. *Inability to handle encrypted traffic*: Traffic to and from SaaS applications are typically encrypted (e.g., transmitted over SSL/TLS). Therefore, even if a traditional DLP solution managed to gain network-level visibility into the traffic, it might not be able to interpret the underlying content.

C. *Interpreting links versus raw data*: Data is never being directly shared in SaaS file sharing applications. Instead what is being shared is some type of link (e.g., a URL) to the content. The link itself reveals little to no useful information about the content being shared. What must be done, therefore, is to analyze the content being pointed to by the link, which is not something that traditional DLP solutions do.

D. *Different sharing semantics*: In the context of traditional enterprise environments, data loss or leakage had a well-defined meaning — namely the crossing of data across the enterprise perimeter. For SaaS file sharing applications, the definition of leakage or loss is fundamentally different as data resides outside the enterprise network. Moreover, it can be shared with third parties who are also outside the network. Traditional DLP solutions do not understand these sharing semantics, and so cannot assess if data is being "lost" or leaked.

E. *Different data model*: Traditional DLP technologies might make different assumptions regarding the data they have to process. For example, they may assume that data is transmitted in a stream and has to be processed as such. When dealing with SaaS based file sharing applications, the data model generally involves being able to access entire files containing sensitive data. Algorithms that are designed for streaming data might not perform well on file-based data (and vice versa). As a result, it is important to develop algorithms that designed to take advantage of full-file content.

F. *Dependence on regular expression and pattern matching*: Traditional DLP technologies rely primarily on basic pattern matching and regular expressions for identifying sensitive content, which can lead to incorrect classification. To address this concern, it is important to apply techniques from natural language processing and machine learning. These approaches go beyond simply trying to understand the raw content, and instead focus on being able to understand the underlying context.

## V. DLP IN CLOUD COMPUTING

Many organizations are moving data to the cloud, but this leads to security and compliance concerns. Though moving to a cloud environment is flexible and cost effective, but the security controls for cloud are very rare.

Having DLP in cloud computing may increase confidence of organizations to move business-critical apps, but this may again lead to questions like how cloud DLP works and how it can actually enhance security and compliance. How it address unique requirements of cloud computing? Data is shifted from central storage form to a distributed model, i.e. from mainframe/midrange to client-server, which forced security organization to change. The risks of data on workstations and in personal devices are directed to an increase in data loss prevention gear, which can monitor mobile and distributed systems. Security management has to discover and track how data is being stored and the new trail of

transmission. Similarly, a shift from physical machines to virtual machines forces another move; the virtual environment introduces many issues, for e.g. security and automation of cloud environments. A thing to be noted is whether cloud providers have the ability to identify sensitive data. Does they have Privacy trade-offs and controls like encryption if they are looking to find the sensitive data on virtual machine to report on. Another thing to be looked into is the granularity in role -based access and reporting. Is the performance impact of finding sensitive data can be managed easily?[3]
In this research paper we will discuss three different approaches of DLP.

### *A. Basic Approach*

In basic approach the following steps to be followed to implement DLP.

1. Data loss prevention (DLP) should find and block the loss of sensitive data. Along with Discovery of sensitive data, a cloud DLP should have the capability to prevent loss of data.
2. Monitor and even block data migrations to and from the cloud from infrastructure. Cloud computing services rely on HTTP as their main communications protocol. Therefore, if HTTP and HTTPS is monitored
3. Finally many potential data migrations across the cloud can be detected [4].
4. The network (SMTP) traffic, along with discovery scans can be sensitized

- by an endpoint agent embedded in the cloud instance
- by routing traffic via a dedicated DLP server or appliance egress to the cloud
- by operating a cloud instance of a DLP server and routing traffic through it

### *B. Extended Approach*

Here basic approach is extended specially to protect regulated and other sensitive information. Translating business policy and rules into a data protection policy creates the following process cycle:

1. *Define*-Create a data protection policy based upon regulatory and business risks.
2. *Detect*- Enable a detection mechanism to identify policy violations.
3. *Enforce*-Determine level of active blocking versus notification or logging based upon the sensitivity of the data and importance of the business activity using the data.

The detection and enforcement processes should be consistently reviewed through dashboard reporting and log files in order to tune the policy definitions. By moving the content inspection point off premise and into the cloud, IT is able to immediately activate a DLP policy that protects the entire enterprise, and sensitive data will be blocked on the first hop into the cloud, before it can fall into the wrong hands. The risks of data breaches spur organizations to perform full inspection of all HTTP and HTTPS traffic leaving the organization, looking for two main categories of violations: ! Regulatory compliance by state or federal governments, or other standards bodies, often pertains to personal or private consumer information. Examples include regulations such as HIPAA, GLBA, PCI, or SOX. ! Company sensitive information may include sales data, pricing information, or intellectual property such as source code. [5]

### *C. Refined Approach*:

Protection of Regulated Information in cloud storage can be provided by an appropriate Data Loss Prevention solution. The steps involved in implementing this protection are described and organized into the following phases:

- *Planning:* The following steps will help the organization make appropriate decisions prior to selecting and implementing a DLP solution to protect information that will be migrated to cloud storage.
- *Assess Current Use of DLP*: Before including DLP in a Cloud strategy, the current use of DLP should be assessed. Ensure that any existing DLP rules may be extended in order to apply the same policy regulations to the cloud data. In some cases it may be desired to apply more stringent controls on data in or intended for cloud storage.
- *Assess Current Use of Cloud Storage:* Similarly, any current use of cloud storage should be understood to determine the protection requirements of the data already stored or to be stored there. It may also be useful, if possible, to understand current cloud use of by employees. It may be found that some enterprise data is already being inappropriately stored in the cloud and creating data loss risks previously not defined.
- *Establish Credible Expectations:* Cloud storage changes the means of visibility and the types of control required over enterprise information. In the absence of a well communicated policy, employees will often use potentially unsecured, cloud services to store confidential data in order to make it more easily accessible from their home or their mobile devices, which may also be unsecured! A DLP solution appropriate for cloud storage protection will apply uniform policy toward information across the enterprise, including cloud storage. In particular, an appropriate DLP solution will provide means for educating end users as well as preventing unauthorized actions when required by policy
- *Set Objectives Appropriate for the Organization:* Gather and review existing policies and procedures concerning the handling of sensitive information. Develop agreement on what information you want to place in cloud storage, what that placement should accomplish, and note any information requiring special protection and control. For example: [6]
    - Records identifying name with SSN.
    - Personal medical or financial records
    - Employees personnel files Cloud storage changes the means of visibility and the types of control required over enterprise information.
- *Involve the Stakeholders*: Ensure the participation of those responsible for managing the use of regulated information and those understanding the regulatory compliance requirements. All parties should understand the benefits

being sought from cloud storage and the requirements for protecting sensitive data expected to be stored there. Managers should understand the benefits and issues of the cloud storage as well as the policy enforcement capabilities provided by DLP.

1. *Migration to the Cloud:* Once the decision is made to proceed with a DLP solution the following steps should be taken to prepare for and execute the migration of information to cloud storage. This refined DLP approach is capable of performing the actions which will help to ensure that regulated information will be properly categorized and protected, or, removed before it may be uploaded and exposed to access in the cloud. There are two basic strategies that may be employed for this migration.

- *Targeted-* A targeted approach employs DLP capabilities to carefully select, assess and, perhaps, remediate specific information assets prior to migrating them to cloud storage. A typical example might be something such as an entire server used by a marketing department that is filled with brochures and other sales collateral. But, with a desire to control any inadvertent release of certain private customer information.
- *Broad-*A broad approach, which may be more common, allows end users to control the migration of their data to a contracted cloud storage provider, but applies DLP to scan and block any regulated data found as it is in flight to the cloud.

In both the approaches DLP Discovery should be employed to inspect all previously stored information in the cloud to bring it under the same policy levels as will be applied to the newly arriving data. These approaches are not mutually exclusive and may be applied at different times with different sets of information, or with different end users.

2. *Operations: By* selecting a DLP solution that provides coverage uniformly across the enterprise including cloud storage, the organization's ongoing management of regulated or other sensitive information is greatly simplified. Policies will be enforced with consistency and from single administrative control. Here are steps to help guide the ongoing processes.

- ▪ *Audit*: A mock compliance audit is conducted which involves the information in Cloud storage. It will force questions to be asked regarding where to focus on risk mitigation strategies.
- ▪ *Scan Large Files Planned for Cloud Storage*: An appropriate DLP solution may be employed to inspect all data poised for sending to the Cloud. Sensitive data discovered will be controlled according to policies established by the enterprise for cloud storage. For efficiency it may sometimes be appropriate to scan entire files when there may be some questions regarding content. Or the files may be large enough that it is desirable to scan them prior to the uploading transmissions which will look at each record at a time.
    - o Before release to the cloud sensitive information may be denied passage or automatically encrypted
    - o Or, other proscribed remediation may be applied
    - o Audit large files with uncertain data content for most efficient handling prior to moving
- *Filter and Audit Information as it is moved to the Cloud*: Apply Network DLP capabilities to inspect all data being sent to the cloud. Before regulated information leaves the network it may be removed, encrypted on the fly or stopped for remediation according to policy for the particular information

Information is inspected at the final stage before leaving the enterprise network
Automatic process reduces opportunities for error
Audit trails provide visibility into information being transmitted
Control is easy to modify if problems are detected

An appropriate DLP solution may be employed to inspect all data poised for sending to the Cloud. Sensitive data discovered will be controlled according to policies established by the enterprise for cloud storage :

*Apply Remediation Selectively at Each Step*: It may or may not be most effective to encrypt-everything sent to the cloud. An appropriate DLP will allows, at every stage in the process, the appropriate remediation to be automatically applied according to the policies established by the enterprise for that particular information and where it is being stored or transmitted

- o Policies dictate action for specific data elements
- o More efficient, speedier processing
- o Alternatives may add burden of needless repetitive encryption and decryption

## VI. DISCUSSIONS

### A. *Advantages of Cloud DLP*

1. 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.
2. The virtual machines can then run client software with a DLP engine that will scan, recognize and block communication of sensitive information.
3. 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.
4. DLP runs as a service, it can be enabled / disabled for virtual machines running in the cloud data center.
5. 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.

6. 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.
7. 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[11].

### *B. Cloud DLP limitations*

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. [2]

## VII. CONCLUSION

Protecting the Benefits Cloud storage provides the enterprise with substantial benefits in cost reductions, scalability, and operational ease. However, as many others have pointed out, the very “sharing” of resources that underlies these advantages must be combined with the proper management of this information. Otherwise new risks of data leakage will be generated. These risks may be deemed a concern if the information being stored is private or sensitive in any way. And, of particularly of concern if it involves data that is regulated by industry or Government rules and laws.

Data Loss Prevention, DLP, technology has proven to be an invaluable resource in protecting regulated data as the enterprise has moved such information from secure data centers to distributed file servers to the desk top and to mobile computing devices. [6]

In our paper we have discussed three different DLP approaches and we have also focused that how DLP has been improved gradually with features to control content in cloud storage. There are many resources to assist organizations in sorting out the options for protection available. But, it is most important to evaluate solutions that will help apply consistent and uniform policy enforcement to information across the entire enterprise, no matter where it is stored, including cloud storage, and that a proof of this capability be demonstrated on site before an organization begins an enterprise implementation.

No single tool is capable of addressing every security issue; however, an appropriate DLP implementation will substantially reduce the risks to an organization as a key component of its overall security strategy.

## REFERENCES

[1] CLOUD COMPUTING PRINCIPLES AND PARADIGMS - by WILLIAM VOORSLUYS, JAMES BROBERG, and RAJKUMAR BUYYA

[2] Richard E. Mackey, Hariharan Sethuraman Mohammed Abdul Haseeb Master (120 credits) Master of Science in Information Security Luleå University of Technology Department of Computer science, Electrical and Space engineering

[3] LTU-EX-2013-41717198.pdf

[4] “Internal report from case”

[5] https://www.zscaler.com/pdf/ebooks/essential_guide_to_cloud_security.pdf

[6] ProtectingDataInTheCloud.pdf DLP Key Features, Available: http://ecs.arrow.com/suppliers/documents/RSA-SolutionBrief-enVisionSolutions.pdf

[7] Webspy, Available:http://www.webspy.com/resources/whitepapers/2008%20WebSpy%20Ltd%20-%20Information%20Security%20and%20Data%20Loss%20Prevention.pdf

[8] Data loss problems, Available: http://www.webspy.com/resources/whitepapers/2009 WebSpy Ltd-Information Security and Data Loss Prevention.pdf

[9] Bradley R. Hunter, Available: http://www.ironport.com/pdf/ironport_dlp_booklet.pdf

[10] David Meizlik, The ROI of Data Loss Prevention, Websense

.