EFFECTIVENESS OF OPEN SOURCE DATA LOSS PREVENTION TOOL IN CLOUD COMPUTING

MAHATHIR BIN SULAIMAN

A project report submitted in partial fulfillment of
the requirement for the award of the degree of
Master of Computer Science (Information Security)

Advanced Informatics School
Universiti Teknologi Malaysia

JUNE 2013

ACKNOWLEDGEMENT

First of all, I would like to express my deepest gratitude to Allah Almighty as with His blessing this project has been successful. I would also like to express my appreciation to my supervisor, Dr Bharanidharan Shanmugam, with his help, guidance and encouragement in preparing this thesis. Lastly, I would also want to extend my appreciation to all of my family members – my wife: Nur Kamariah, my children: Sakina, Mahfudz, Safiyyah and Muzaffar and not forgotten my manager, colleagues and others who have provided assistance and support directly and indirectly.

ABSTRACT

With the systems and applications migration from the traditional enterprise data center infrastructures to the virtualisation of cloud computing infrastructure, there are changes required in term of the way of system security and data security management. There are more various sources of threats to the data integrity preservation which may come from internal employees, external users, cloud providers and the vendor of cloud providers. Among the way to ensure the data is not leaked out is by looking at data loss prevention tool. The tool should be protecting data at all the three common states namely data in motion, data in transmit and data at rest. The thesis intention is to find out the effectiveness of using open source data loss prevention in cloud computing infrastructure. In addition, the thesis would also study the security vulnerabilities on the open source DLP deployment architecture system and propose the improved architecture setup. While the effectiveness evaluation is done using open source, there is also a need to find the market leading commercial data loss prevention tool and identified the market strength.

ABSTRAK

Pemindahan sistem dan aplikasi dari tradisional pengkomputeran infrastruktur kepada infrastruktur pengkomputeran awan, beberapa perubahan yang diperlukan dalam pengurusan system keselamatan dan pengurusan data. Terdapat bermacam sumber ancaman serangan keselamatan yang lebih khusus kepada pemeliharaan integriti data yang boleh datang dari kakitangan organisasi sendiri, pengguna luar, pekerja sokongan perkhidmatan perkomputeran awan dan pekerja sokongan pembekal pekakas perkhidmatan awan. Di antara cara untuk memastikan data tidak bocor keluar adalah dengan perlaksanaan perisian pencegahan kehilangan data (DLP). Perisian ini perlu berkebolehan untuk melindungi data pada kesemua ketiga-tiga keadaan data iaitu data ketika digunakan, data ketika dalam penghantaran dan data ketika disimpan. Sejerus itu, tujuan tesis adalah untuk mengetahui keberkesanan penggunaan perisian pencegahan kehilangan data, dari jenis sumber terbuka, dalam infrastruktur perkomputeran awan. Di samping itu, tesis ini juga akan mengkaji kelemahan keselamatan pada sumber terbuka DLP dari segi perlaksanaan pemasangan seni bina dan seterusnya mencadangkan rancangan persediaan yang lebih baik. Walaupun penilaian keberkesanan dilakukan dengan menggunakan sumber terbuka DLP, terdapat juga keperluan untuk mengenalpasti peneraju pasaran komersil untuk perisian pencegahan kehilangan data dan mendalami kekuatan perisian tersebut di pasaran.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABBREVIATIONS		DESCRIPTION
DLP	-	Data Loss Prevention
OS	-	Operating System
DOS	-	Disk Operating System
IT	-	Information Technology
SP	-	Service Provider
HTTP	-	Hyper Text Transfer Protocol
SMTP	-	Send Mail Transfer Protocol
CMF	-	Content Monitoring and Filtering

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1. Background

Cloud computing, an emerging information technology approach was a transformed from grid computing architecture. The cloud computing terminology, as defined by NIST is "*a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort*" (Mell and Grance, 2011). This new information technology (IT) infrastructure model is positioned as a cost reduction approach in IT operation and acquisition of server, storage and data center for non IT business organization. The primary benefit for customers of cloud computing infrastructure is that they would be able to focus more on core business operations and will not have to deal with the high cost of IT operations. Their IT cost would be minimized to subscription cost for the IT operations to the cloud computing service provider (SP) and save operation cost on the maintenance of the data center operation, servers, storage and network cost.

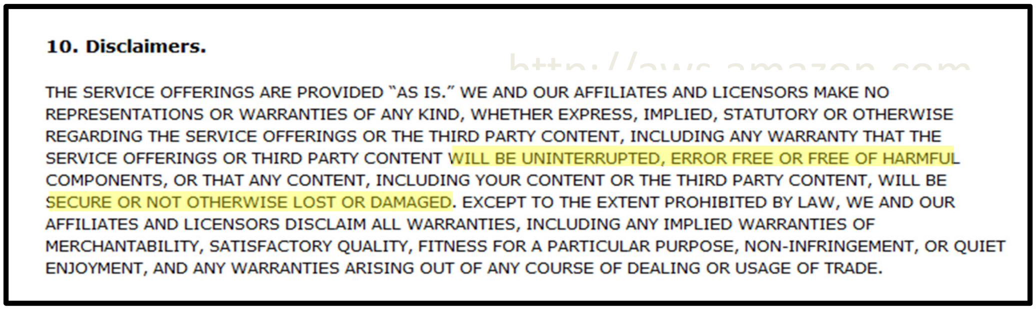
In addition, their internal IT support teams within the business may be reduced. As a result of that, the business owners could hire personnel for their other core business operations such as sales and marketing officers. However, cloud computing is still a new technology field compared to other technology fields. There are few major security issues and processes that need to be addressed by cloud computing SP in order to attract customers. According to a survey by IDC in 2008, which was conducted on 244 IT executives/CIOs and their line-of-business colleagues about their companies' use of, and views about IT Cloud Services; the finding was 74.6% of IT managers and CIOs believed that the primary challenge that hinders them from using cloud computing services is the security issue on cloud computing (IDC, 2008). The cloud service providers struggle with the cloud environment security issues because the cloud model is very complicated and has a many dimensions that must be evaluated when establishing a comprehensive security model (Almorsy et al, 2010).

One of the cloud computing security concerns is on the data integrity privacy. This is due to the technology complexity on the cloud computing. According to Subashini and Kavitha (2010), there are three types of cloud computing delivery models which are the Service as a Server (SaaS), Platform as a Service (PaaS) and Infrastructure as a service (IaaS). These service models have a different level of security requirement in the cloud environment. IaaS is the under-laying foundation of all cloud services, with PaaS built upon it and SaaS in turn built upon it. Just as capabilities are inherited, so are the information security issues and risks.

1.2. Background of the Problem

The data integrity preservation is the defense of true state of data that are being in attributes of completeness, accurate and not being accessed by unauthorized party for read or write capability (Boritz, 2011). Since the data is not stored on the customers' premises and are not physically isolated from other organisations, there would be some level of uncertainty of securing the data integrity satisfactorily. Traditionally, customers would have total control on the infrastructure and their own premise physical security would contribute to their higher confident level. Due to the nature of multi-tenancy in cloud computing, service providers (SP) will not share the detail infrastructure information with their customers. Multi-tenancy which results in virtualizing the boundaries among the hosted application services of different customers and tenants and thus the cloud platform need to security harden such boundaries with new category of security controls (AlMorsy et al., 2011). This is to protect the security of the data center and part of physical security strategy.

Secondly, major cloud computing providers do actually state in their agreement that they had no warranties to ensure data preservation. They have clearly stated in their Service Level Agreement that they would not be held responsible for any security incidents and information leakage (Hoffman, 2012). For example, Amazon in its offering of cloud computing services, mentioned in end user level agreement (EULA) (as per Figure 1 below), among others means that their services are provided as is and that data hosted provided by the customers or third party vendor is not error free and may be damage or loss.



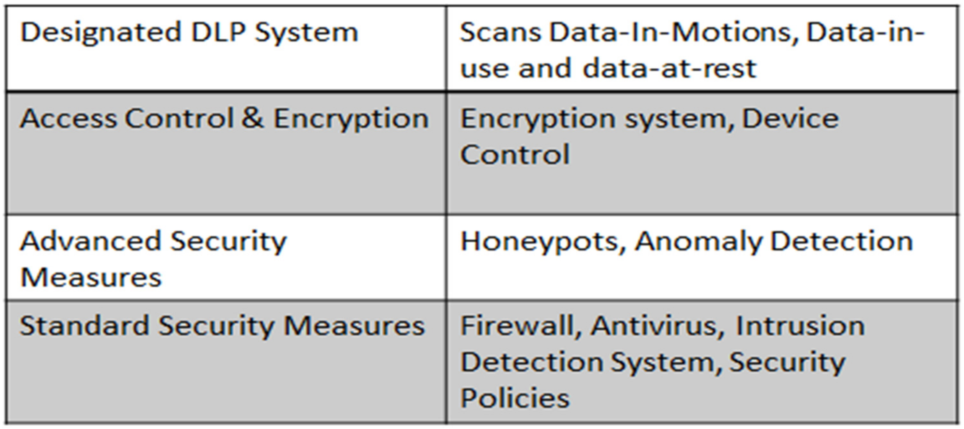
10. Disclaimers.

THE SERVICE OFFERINGS ARE PROVIDED "AS IS." WE AND OUR AFFILIATES AND LICENSORS MAKE NO REPRESENTATIONS OR WARRANTIES OF ANY KIND, WHETHER EXPRESS, IMPLIED, STATUTORY OR OTHERWISE REGARDING THE SERVICE OFFERINGS OR THE THIRD PARTY CONTENT, INCLUDING ANY WARRANTY THAT THE SERVICE OFFERINGS OR THIRD PARTY CONTENT WILL BE UNINTERRUPTED, ERROR FREE OR FREE OF HARMFUL COMPONENTS, OR THAT ANY CONTENT, INCLUDING YOUR CONTENT OR THE THIRD PARTY CONTENT, WILL BE SECURE OR NOT OTHERWISE LOST OR DAMAGED. EXCEPT TO THE EXTENT PROHIBITED BY LAW, WE AND OUR AFFILIATES AND LICENSORS DISCLAIM ALL WARRANTIES, INCLUDING ANY IMPLIED WARRANTIES OF MERCHANTABILITY, SATISFACTORY QUALITY, FITNESS FOR A PARTICULAR PURPOSE, NON-INFRINGEMENT, OR QUIET ENJOYMENT, AND ANY WARRANTIES ARISING OUT OF ANY COURSE OF DEALING OR USAGE OF TRADE.

Figure 1.1: Snapshot of Amazon Cloud End User Level Agreement

Because of these concerns, there had been various studies and researches that are looking into the area to ensure the interest of consumers are protected. Shabtai mentioned as per in Figure 1.2 that there are several techniques of data loss prevention including designated DLP system, access control, advanced and standard security measures (2012). Ristenpart *et al.* (2009) showcased that Amazon cloud is prone to side-channel attacks and it would be possible to capture and steal data, once the malicious virtual machine is placed on the same server as its target. It is possible to carefully monitor how access to resources fluctuates and thereby potentially glean sensitive information about a victim. However, they acknowledge, that there are a number of factors that would make such an attack significantly more difficult in practice.

Typical security measures in guarding the data leakage, SP is putting up protection tools as per Figure 1.2. Starting from standard security measures which are network firewall, intrusion detection system and other network related security devices, this layer help to mitigate the network layer attacks. But as the attack had moved up to the application layer, DLP is needed as the security assurance, in case the bottom level were compromised.



Designated DLP System	Scans Data-In-Motions, Data-in-use and data-at-rest
Access Control & Encryption	Encryption system, Device Control
Advanced Security Measures	Honeypots, Anomaly Detection
Standard Security Measures	Firewall, Antivirus, Intrusion Detection System, Security Policies

Figure 1.2: Technological Approaches on Data Leakage

1.3. Problem Statement

Due to the nature of cloud computing, the data stored in cloud computing environment are exposed to the risk of data leakage resulted from unauthorized data sharing, unauthorized data access and unauthorized data modification. In the cloud computing infrastructure, users of cloud services have serious threat of losing confidential data. Firstly, cloud computing is a multi-tenancy infrastructure which means the data stored could be sitting next to business competitors' data. Secondly, cloud computing data is based on distributed data structure system where the data is hosted in unknown location which made it more worrisome. Moreover, the data owner does not have any visibility on SP system administrator's activities. Even though, they are bound by policy, the fact that human weakness for the corruption could be weakness that worried customers. Based on preliminary research finding, there is few studies was done in preserving the data integrity on the cloud using data loss protection (DLP) software tool.

1.4. Research Questions

Data loss prevention (DLP) tool is the software of preventing sensitive data from leaving a user's device to the unauthorised destination (Liu, 2010). The objective of this paper would be leading to providing the solution of ensuring the data integrity even though they are hosted on the cloud computing infrastructure. Acknowledging the critical of data security, there had been studies in the long term that there are cloud infrastructure can be implemented with the secured cloud in the design as per discussed by Shaikh and Haider (2011). This thesis would study the functional metric comparison of the existing data loss protection tools and to elaborate the effectiveness of DLP implementation in preserving data integrity in cloud computing. The research questions are as follows:

1) What are the available data loss prevention solutions available in the market and which one is market leader in the industry?

2) How effective open source DLP in IaaS type of cloud computing infrastructure in preventing data loss?

3) What are security weaknesses in the standard open source DLP system architecture deployment and proposed improved DLP system architecture?

1.5. Research Objectives

1) To examine the available data loss prevention solutions available in the market and identify market leader in the industry.

2) To evaluate on the effectiveness of open source DLP in IaaS type of cloud computing infrastructure.

3) To identify security weaknesses in the standard open source DLP system architecture deployment and propose an improved DLP system architecture.

1.6. Project Aim

The aim of this project is address to the effectiveness of the data integrity preservation in cloud computing infrastructure by using DLP software tool and analyse the functionality and features for major DLP software in the market. We are targeting to complete the following:

1) Conclude the investigation on data integrity preservation can be compromised in cloud computing without the presence of data loss protection tool.

2) Identify and produce comparison metric for the available data loss protection tools in the market that can be deployed in cloud computing environment.

3) Conclude the effectiveness of DLP tool in preserving the data integrity in cloud computing.

1.7. Project Scope

The thesis will be using both the experimental research methodology and the latest available research data on data loss prevention. There are three types of cloud computing infrastructures: Platform as a Service (PaaS), Software as a Service (SaaS) and Infrastructure as a Service (IaaS). For this thesis, we will only be covering for the type of Infrastructure as Services (IaaS) due to the only platform that the customers can have installation and configuration done on the cloud and have full control of the server. For this research purpose, there will be the setting up of cloud computing using open source virtualization software of Oracle VM VirtualBox. Once the mini cloud is ready, there will be proof of concept of to verify the data stored in the cloud are exposed to the risk of unauthorized access and evaluate the standard DLP system architecture is sufficiently secured and reliable. The project would focus on setting up of open source DLP tool deployment and configuration and demonstrate the potential data risk. The verification of proposed DLP tool would be done on the same cloud architecture and prove the data is protected even though it is hosted on the cloud computing infrastructure which the data center can be operated from foreign countries.

1.8. Summary

In summary, cloud computing infrastructure offers the business organisations opportunity to reduce their respective IT cost. However, with this change in business approach in term of IT strategy, there is a need to diligently manage the system security and data protection. The fact that the system and their data are out from their physical premise and being managed by the cloud provider who is potentially may also run their business competitors system and data, the business organisation should be more aggressive in protecting their data. On top of that, there is no control or potential audit can be done by the users of cloud computing to ensure there is no off-the-record activities by the support personnel of both cloud computing providers and the their hardware and software vendor. To reduce the risk of the data leakage is by using the DLP tool for all system and applications deployed in the cloud computing environment.

REFERENCES

A. Shabtai et al., (2012) *A Survey of Data Leakage Detection and Prevention Solutions*, SpringerBriefs in Computer Science. 20 – 25

Baliga, J; Ayre,R; Hinton,K; Tucker, R. (January 2011). *Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport*. Proceedings of the IEEE. Vol. 99(1).151

Boritz, J. (Aug 2011). *I'S Practitioners' Views on Core Concepts of Information Integrity*". International Journal of Accounting Information Systems.

Bosworth, M. (2008) *ChoicePoint Settles Data Breach Lawsuit*. ConsumerAffairs.Com. 27 January 2008. [online]. Retrieved Feb 15, 2013 http://www.consumeraffairs.com/news04/2008/01/choicepoint_settle.html

Chen, D; Zhao H. (2012) *Data Security and Privacy Protection Issues in Cloud Computing*. 2012 International Conference on Computer Science and Electronics Engineering. 647-651

Gartner (2013). *Magic Quadrants and Market Scopes: How Gartner Evaluates Vendors Within a Market.* Gartner. [online]. Retrieved Feb 20, 2013. http://www.gartner.com/DisplayDocument?doc_cd=131166

Gessiou, Eleni, Vu, Quang Hieu and Ioannidis, Sotiris(2010). *IRILD: an Information Retrieval based method for Information Leak Detection.* Institute of Computer Science, FORTH, Greece and Etisalat BT Innovation Center, Khalifa University, UAE.

Hoffman, M. (2012) *Cloud Computing: The Next Headache*. Cloud Computing Security Seminar at SKMM Office, Cyberjaya.

Hart, Michael, Manadhata, Pratyusa and Johnson, Rob. *Text Classification for Data Loss Prevention.* s.l. : Springer Berlin / Heidelberg, 2011. Vol. 6794, pp. 18-37.

IDC, *IT Cloud Services User Survey, pt.2: Top Benefits & Challenges* [Online]. Retrieved on Dec 10th ,2012 at http://blogs.idc.com/ie/?p=210

Info-Tech (2012). Vendor Landscape Storyboard: Data Loss Prevention [Online]. Retrieved on March 15, 2012. From http://www.infotech.com/research/it-storyboard-select-an-enterprise-data-loss-prevention-solution

J. Albuquerque, H. Krumm and P. de Geus,(2008) "*Model-based management of security services in complex network environments*," *IEEE Network Operations and Management Symposiu*, pp. 1031-1036, Salvador.

Kanagasingham,P (Aug 15, 2008). *Data Loss Prevention*. SANS Institute InfoSec Reading Room. [Online]. Retrieved on March 15, 2012.from http://www.sans.org/ reading_room/ whitepapers/ dlp/ data-loss-prevention_32883

Kandukuri BR, Paturi VR, Rakshit A. (2009) *"Cloud security issues"*. IEEE international conference on services computing, 517–20.

Keila, P.S. and Skillicorn, D.B.(2005) *Detecting Unusual Email Communication..* 2005 conference of the Centre for Advanced Studies on Collaborative research

Liu, S; Kuhn,R. (2010) D*ata Loss Prevention*, IT Professional Vol. 12(2), p10–13.

M. Almorsy, J. Grundy, I. Mueller,(2010) *An analysis of the cloud computing security problem,* 2010 Asia Pacific Cloud Workshop Australia.

M. Almorsy, J. Grundy, S. Amani, (2011) *Collaboration-Based Cloud Computing Security Management Framework.* 2011 IEEE 4th International Conference on Cloud Computing. 364-371

Ranchal, R.; Bhargava, B.; Othmane, L.B.; Lilien, L.; Anya Kim; Myong Kang; Linderman, M.; (2010) *Protection of Identity Information in Cloud.* 29th IEEE International Symposium on Reliable Distributed Systems. Vol 29, 368 - 372

Mell,P; Grance, T. (2011) *The NIST Working Definition of Cloud Computing* v14, Nat.Inst. Standards Technology, [Online]. Retrieved on Mar 15, 2012 from: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf

Mishra, R.; Dash, S.K.; D.P.; Tripathy, A. (2011). *A Privacy Preserving Repository for Securing Data across the Cloud.* Electronics Computer Technology (ICECT), 3rd International Conference:6-10

Miranda & Siani,(2009).*A Client-Based Privacy Manager for Cloud Computing*, COMSWARE'09, Dublin, Ireland

Mogull, Rich. *Best Practices for Endpoint Data Loss Prevention.* Securosis, L.L.C., 2009.

Manuel, Stephane (2008) *Classification and Generation of Disturbance Vector for Collision Attacks Against SHA-1. iacr.org.* [Online]. Retrieved Feb 15, 2013 http://eprint.iacr.org/2008/469.pdf.

Polatcan, Onur, Mishra, Sumita and Pan, Yin (2011) New York : *E-mail Behavior Profiling based on Attachment Type and Language*, 6th Annual Symposium on Information Assurance (ASIA '11). 6-10.

Phua C,(2009), *Protecting organisations from personal data breaches.* Computer Fraud. p15-17

R. Gellman (2009), *Privacy in the Clouds: Risks to Privacy and Confidentiality from Cloud Computing,* World Privacy Forum, Feb. 2009. [Online] http://www.worldprivacyforum.org/pdf/WPF_Cloud_Privacy_Report.pdf

Rongxing et al, (2010) ”Secure Provenance: The Essential Bread and Butter of Data Forensics in Cloud Computing”, ASIACCS '10 Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security, Beijing, China.

Reddy, VK and Reedy, L.S.S. (September 2011). *Security Architecture of Cloud Computing*. International Journal of Engineering Science and Technology (IJEST). Vol. 3(9)

Rohit,R; Bharat, B; Othmane, L; Leszek,L;(2010) *Protection of Identity Information in Cloud Computing without Trusted Third Party.* 29th IEEE International Symposium on Reliable Distributed Systems:389

Subashini S , Kavitha V, (2010) *A survey on security issues in service delivery models of cloud computing*. Journal of Network and Computer Applications.

Shaikh, F.B. and Haider, S. (December 2011) *Security Threats in Cloud Computing*. 6th International Conference on Internet Technology and Secured Transactions, Abu Dhabi, United Arab Emirates.

SUN Microsystems (2010) *Sun Cloud Architecture Introduction White Paper.* [Online]. Retrieved on April 1, 2012 from http://developers.sun.com.cn/blog/ functionalca/ resource/ sun_353cloudcomputing_chchine.pdf.

T. Ristenpart, E. Tromer, H. Shacham, S. Savage (2009) *Hey, You, Get Off My Cloud: Exploring Information Leakage in Third-Party Compute Clouds. 6th ACM conference on Computer and Communications Security*, Chicago. 199-212.

Trend Micro (2010). *Trend Micro DLP Administrator's Guide.* s.l. : Trend Micro.

Wenchao, Z; Sherr, M; Marczak, W; Zhang, Z; Tao, T; Loo, BT; Lee, I (2010) *Towards a Data-centric View of Cloud Security,* CloudDB 2010, Toronto, Canada