International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-3, June2015
ISSN: 2395-3470
www.ijseas.com

Detecting Data Leakage in Cloud Computing Environment (A Case Study of General Hospital Software)

Alex Ofori Karikari[1], Joseph Kobina Panford[2], James Ben Hayfron-Acquah[3], Frimpong Twum[4]

[1, 2, 3, 4] Department of Computer Science, KNUST, Kumasi, Ghana

Abstract

Generally sensitive data are leaked by users or data entry operators and identifying those operators is paramount and should be done at an early stage to forestall any catastrophe. In distributed networks, data should by all means be handed over to data operators or trusted third parties. Computing systems have been infiltrated by persons to cause irreversible damages to institutions and organizations worldwide and employee factor in data leakages/breaches has been on the ascendancy, through intentional or unintentional leakages/breaches. So on this premise the research seeks to use user accessibility system which will cure non-repudiation and Transaction log /Audit Trail monitoring system to see whether one or more agents can be detected to have leaked data. Analysis of the old hospital system design was conducted and the finding was that, the system does not have any Transaction log / Audit Trail system so anything entered or deleted from the system, for example, could not be tracked by the system administrator. So based on the analysis, a new system was designed which has Transaction log / Audit Trail system in place so all activities either within or outside the hospital are monitored and has a date and Time Stamp. So in conclusion, the new system gives the administrator the chance to monitor all activities in the system.

***Keywords**: Data Leakage, Audit Trail, Cloud Computing, SOA, Middleware, Virtualization*

1. Introduction

This section gives the overview of the topic and establishes understanding of what Data Leakage is and the importance of detecting the leakage and plugging it. It also explains the types of data that exists, that is, Data at Rest, Data in Use and Data in Motion. The writers went further to establish that all these types of data are vulnerable to leakages. The types of data leakages or loss as being Intentional, Unintentional, Disaster, Failure and Crime are also discussed. Because the research is aimed at detecting intentional data leakages from an inside agent, types of intentional exploitations are illustrated.
Some of the examples are sited as Malicious Insider, Malware, Code Infections, and Phishing.

1.1 Problem Statement

The problem statement established the fact that data leakages are a big concern to organizations as well as individuals and also costing huge sums of money to institutions. It is also buttressed by INFOWATCH's "Global Data Leakage Report 2009" which states that, 51% of data leakages were resulted from intentional attacks and 43% of the leakages were due to accidental events, which indicates a strong increase on intentional leakages as comparing to 2007 figures (i.e. 29% intentional and 71% accidental).

1.2 Research Question

Data leakage which by definition is uncontrolled, unauthorized transmission of classified information from a Data Centre or computer system to the outside is causing companies or organizations huge sums of money and also breach of trusts.

Now individuals as well as the corporate bodies are moving to the cloud and as such data leakages have become a challenge.

How can data leakages be detected especially through user activities before it gets out of the institution?

1.3 Objectives

The main objective of this research is to develop an application which profiles user access activities into a computing resource and tracks user activities to detect possible data breaches and profile leaking data source.

1.4 Significance of Study

Data leakage happens every day where confidential information such as customer or patient data, source code or design specification, price hits, intellectual property, trade secrets, etc. are leaked out. When these are leaked it leaves the company unprotected and goes outside the jurisdiction of the cooperation. This uncontrolled data puts business in a vulnerable position. Once this data is no longer within the domain, then the company is at serious risk. This research seeks to help detect leakage and sensitive data in a cloud environment because when cybercriminals, for instance, sell this data for profit, it cost organizations money, damages.

1.5 Scope of Research

This research is intended to show how user activities can be monitored, through system transaction logs to determine data leakage channels and by whom, through the development of an application that tracks user activities through user accessibility to a computing resource. Additionally, included will be a transaction log which will profile user activities such as login dates and activities.

2. Literature Review

This literature review will talk about the works of the other writers in relation to data leakages and cloud computing.

The chapter will also look at the security aspect of the cloud computing and the types of cloud services that are rendered by venders.

This literature will touch on how cloud computing evolved and take closer look at the architecture of the cloud computing because of the motive of using cloud computing to detect data leakages.

It will be very necessary to get a fundamental overview of the development of the distributed computing, its first occurrence and how it evolved. It is also paramount to clearly clarify what cloud computing is, which concepts it involved.

In 2008, they regarded cloud computing as a "collection of many old and few concept in several research fields like Service- Oriented Architectures (SOA), distributed and grid computing as well as virtualization". According to them "cloud computing can be considered as a new computing paradigm that allows users to temporary utilize computing infrastructure over the network, supplied as a service by the cloud provider at possibly one or more levels of abstraction "(Youseff et al . 2008).

According to Armbrust et al (2009), "Cloud computing refers to both the application delivered as services over the internet and the hardware and systems software in the data centers that provide those services. The services themselves have long been referred to as Software –as-a-Service (SaaS). The data centre hardware and software is what we will call a cloud. When a cloud is made available in pay- as- you- go manner to the general public, we call it a public cloud: the service being sold is utility computing. We use the term private cloud to refer to internal data centers of a business or other organization, not made available to the general public. Thus cloud computing is the sum of SaaS and utility computing, but does not include private clouds".

Consulting firms, market research companies and cloud computing service providers are originators of most the cloud computing definitions. The market research company IDC (International Data Corporations) defines cloud computing in general as " an emerging IT development, deployment and delivery model, enabling real time delivery of products, services and solutions over its internet"(Gens 2008). By this definition, it means cloud computing is the technical basis for cloud services, offering consumer and business solution that are consumed in real-time over the internet. The technological foundation of the cloud computing includes infrastructure, system software, application development and deployment software, system and application management software as well as IP based network services.

2.1 Data leakage

According to Miller (2009) , there has been a shift from when organizations only battle to be free from intrusions, viruses, and spam, and are now wrestling to be free from Data Leakage , be it intentional or accidental exposure of information ranging from legally protected personal information to intellectual property and trade secrets. The exposure of confidential information has become the number one threat for many organizations these days.

Davis (2009) described data leakage as when classified information is transferred illegally to the outside world.

3. Methodology

The old Hospital software system which does not have any form of checks for what users do in the system was compared with the proposed system which has a transaction log /Audit Trail to track all user transactions and also few of the users were involved in an oral interview.

The main purpose of the study was to show how user accessibility and activity to a computing resource could be tracked, monitored and audited to safeguard the integrity, accessibility and availability of data to authorized and authenticated users in cloud applications. The strategy was to adopt *Normal Hospital Software* as a case study. This is because these hospital software hosted in cloud environments have no local databases, different access levels and remote user authentication and authorizations.

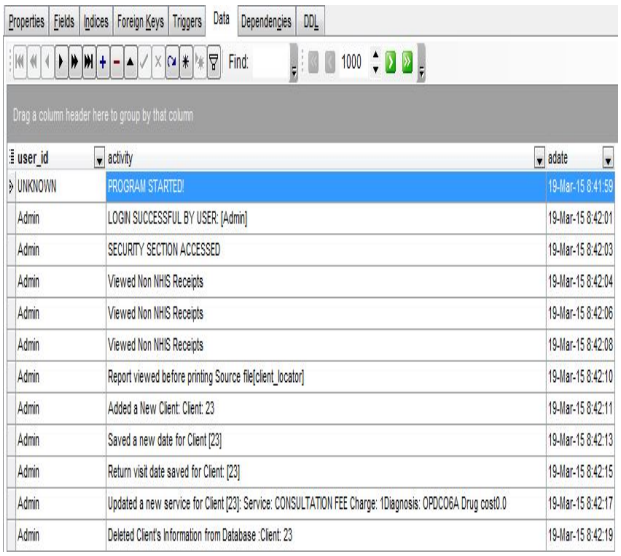
3.1 Analysis and Design

The system was seen not to have any audit trailing process and procedure considering

the nature and environment within which it operates (cloud computing environment). Another area of interest to this study was an audit and monitoring trail measure, it was shown not to have any system in place that audits and monitors system users and this meant that user activities could not be traced or monitored for auditing and security assessment purposes through the use of a transaction log monitor.

3.2 The Proposed System

The New application system has in it an audit trail/transaction log system which monitors user activities to computing resource at every stage of the process.



user_id	activity	adate
UNKNOWN	PROGRAM STARTED!	19-Mar-15 8:41:59
Admin	LOGIN SUCCESSFUL BY USER: [Admin]	19-Mar-15 8:42:01
Admin	SECURITY SECTION ACCESSED	19-Mar-15 8:42:03
Admin	Viewed Non NHIS Receipts	19-Mar-15 8:42:04
Admin	Viewed Non NHIS Receipts	19-Mar-15 8:42:06
Admin	Viewed Non NHIS Receipts	19-Mar-15 8:42:08
Admin	Report viewed before printing Source file[client_locator]	19-Mar-15 8:42:10
Admin	Added a New Client: Client: 23	19-Mar-15 8:42:11
Admin	Saved a new date for Client [23]	19-Mar-15 8:42:13
Admin	Return visit date saved for Client: [23]	19-Mar-15 8:42:15
Admin	Updated a new service for Client [23]: Service: CONSULTATION FEE Charge: 1Diagnosis: OPDC06A Drug cost0.0	19-Mar-15 8:42:17
Admin	Deleted Client's Information from Database :Client: 23	19-Mar-15 8:42:19

Figure 1: Transaction log/Audit Trail

Figure 1 hows the various activities performed by the user (admin) including the date and time and actual activity performed at each stage of the application process. This audit trail table has no link or dependency to the other tables in the program making it safe from manipulations from other tables. It is intended to be activated at every login stage for user identification and verification purposes. It has been tied to application accessibility, which means a user must be successfully verified before access is granted into any software application. This will ensure that users who are not biometrically verified through this system have no access to any software program. It is just another level of the user identification and authorization within the cloud computing architecture (a middleware application).

With the adoption of cloud computing services by many institutions it is prudent to have a system that will monitor all activities of users so that even when an intruder with different level of authorization or not recognized name in the organization gain access it can be recognized by the Administrator and quickly find ways of resolving the issue. In hospitals, information pertaining to patient are very important so any leakage of information to an outsider is dangerous to the hospital and also can be used to demand ransom from patients, especially prominent people in society with the treat of releasing those health related information.

3.3 Simulation-Network Design Topology

Figure 2, depicts how the simulation test was done with the dual purpose of how a wrongly configured network topology could grant access to both known and unknown user (data clerk) into the system to breach/leak data for instance.

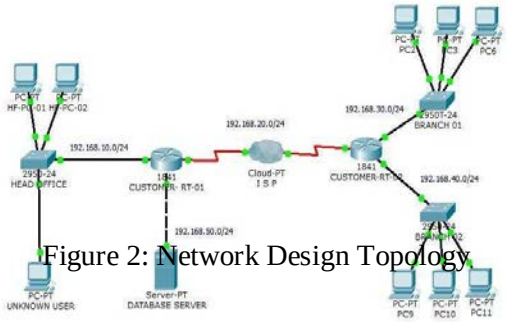


Figure 2: Network Design Topology

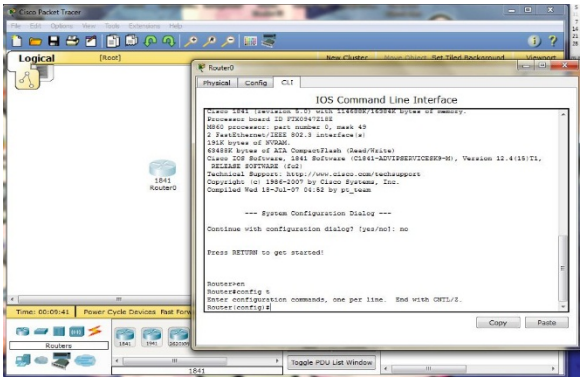


Figure 3: Configuration interface of a router

Figure 3, shows the configuration interface of a Cisco with a router for configuration settings. The Command Line Interface (CLI) is the configuration interface where all security settings and other configuration are done.

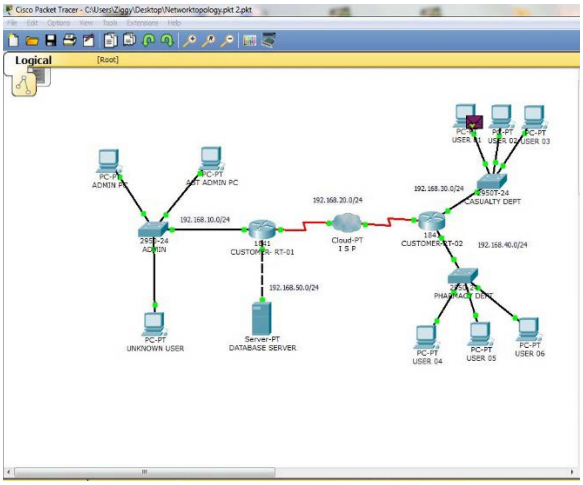


Figure 4: Network topology diagram

Figure 4, is a prototype of how the network application is implemented at the case study sites. Workstations were installed at the various departments (e.g. Out Patients Department, Casualty, pharmacy etc.). The workstations are configured to allow a specific number of data entry clerks through Virtual Local Area Network (VLan) allocations. VLan usage ensures that network infrastructure is used effectively.

The link to the head office was done by an Internet Service Provider (ISP).

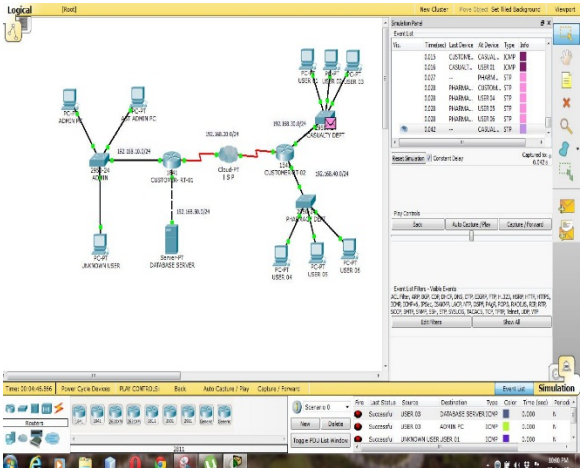


Figure 5: Positive packet sent/received responses

Figure 5, depicts how packet/data was received successfully from the Administrator's PC to the USER 03 in the casualty Department as shown in the event lists.

3.4 Evaluation

From the analysis it can be inferred that all the stated research objectives have been achieved. This research project has been successfully created to show through a transaction log profile how data could be leaked and user's activities to identify who leaks data within a specified database environment. The system now has the capability to authorized users, authenticate users, capture user activities in a form of transaction log and finally manage user access into a software program based on verification process. With the transaction log, the issue of non-repudiation is resolved as a user's logon activities are captured at all times. Additionally, all database entries, that is, user passwords are encrypted in the database system.

4. Conclusions

Data leakages cost organizations/institutions; higher cost in

terms of money and also reputation, so it is prudent to have a system whereby system administrators will monitor all activities of users logged on into the system. So this research in a way will help system administrators track down at least users in their system.

References

Ajinkya S. Yardav, Ravinda P. Bachate and Shadab A. Pattekari (Prof) "Detection of Data leakage Using Unobstructive Techniques" Journal of Computer Engineering (108R – JCE, Volume 8 Issue 4 (Jan – Feb 2013), pp27 – 84.

Ambrust Michael, Amando Fox, Rean Griffith, Anthony D, Joseph, Randy Kats, Andy Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica and Mati Zaharia (2009) : Above the Clouds: A Berkeley view of cloud computing , p.1.

Bill Claybrook (2010): Cloud vs. In-house: Where to run that App? Computer World Journal, Page 1.

Dahal Sanjaya (2012), Security Architecture for Cloud computing Platform, pp24-26.

Davis Ziff (2009) PC Magazine Encyclopaedia, Definition of Cloud.

Fowler A. Geoffrey and Ben Worthen (2009): The Internet industry Is on a Cloud – Whatever That May Mean. (ww.wsj.co/articles/SB123802623665542725, Date assessed: September, 2013.)

Infowatch "Global Data Leakage Report, 2009 (www.infowatch.com, Date assessed: September, 2013)

Shaw Jack: Dynasis Blue Paper: Cloud computing Public, Private and Hybrid (www.Dynasis.com, Date assessed Nov., 2013)

Stedum James, A Brief History of Cloud Computing, July 29, 2013 (Posted on SoftLayer Blog , Pp. 1)

Larry Ellision, CEO of Oracle (2007): Analysis Conference in 2007.

Lawrence C. Miller (2009), Data Leakage for Dummies, Pp. 1-10.

Markus Bohm, Stefanie Leimeister, Christoph Riedl,Helmut Kremar of Technische Universitat Munchen (Tum) "Cloud Computing and Computing Evolution", Pp. 6.

Mell Peter and Timothy Grance (2011), The National Institute of Standard and Technology (NIST) Publication (Pp. 800 – 145): Definition of Cloud Computing.

Peter Gordon (2007), Data Leakage – Threats and Mitigation, Pp. 5-6.

Plummer,D., Clearly,D., and Smith , D. (2008) Cloud Computing – Confusion leads to Opportunity, Pp. 10-20.

Schneier , Bruce (2010): Data at Rest Vs. Data In Motion, Pp. 1.

Sean Carlin and Kelvin Curran, University of Ulster, UK. "Cloud computing security", International Journal Ambient Computing and Intelligence, Volume 3 issue 1 Pp. 1-6.

Singleton, Tommie, (2010): The Minimum IT Controls to Assess in a Financial Audit (Part II), ISACA Journal, vol. 2, Issue. 3

Sun Microsystems 2009, Cloud Computing architecture.

Tim Mather, Subra Kumaraswamy, and Shahed Latif (2009): Cloud Security and Privacy- an enterprise perspective on risks and compliance, Pp. 1-30.

Vasquero Luis M., Luis Rodero-Merino, Juan Caceres, Maik Lindner (2009): A Break in the Clouds, towards a Cloud Definition, Pp. 51.

Youseff, L, Butrico, M, and Da Silva, D: Toward a Unified Ontology of Cloud Computing, Pp 1-2.