

Optimized Method for Tracking of Data Leaving the Cloud Environment

Lokendra Vishwakarma ^{#1}, Pallavi Gupta ^{#2}, Awadheshwari patel ^{#3}

^{#1, 2, 3} *Computer Science and Engineering Department*
^{#1, 2, 3} *Samrat Ashok Technological Institute, Vidisha(MP), India.*

Abstract- Information security in the digital world is the biggest concern in these days. And as the cloud computing is also a part of this digital computing so these concerns also adhered with this new era of technology. As the Data and Information leakages out of the cloud computing environments are the fundamental cloud security concerns for both cloud consumer and cloud service providers. A literature survey of this new technology revealed the security issues of current technology and need for a new methodology. In this paper we discuss the requirements and proposes a novel auditing methodology that enables tracking of data transferred out of the Clouds and if something has wrong happened so will be able to retrieve the previous status of their data so in that case data and information is secured at both ends i.e. consumer and service provider. This research get benefited to the owner of the data so that they sure to keep their data into the cloud.

Keywords- Cloud computing Security Challenges, Data tracking, Cryptography.

1. INTRODUCTION

Cloud computing has been defined by NIST as a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or cloud provider interaction[2]. Cloud computing is a large-scale distributed computing paradigm driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, highly available and configurable and reconfigurable computing resources can be rapidly provisioned and released with minimal management effort in the data centers. Services are delivered on demand to external customers over high-speed Internet with the “X as a service (XasS)” computing architecture, which is broken down into three segments: “applications”, “platforms”, and “infrastructure”[18].

Cloud computing is all about delivering Infrastructure, platform and software as a service which is reliable, scalable and economical for hosting web based applications.

It combines concepts of distributed, grid and utility computing and basically deals with resource allocation and service provisioning to meet dynamically changing needs [14]. Main objective of cloud computing is to enable datacenters to meet out users expectations and allow them to access and deploy applications from any corner of world with better QoS (Quality of Service) [15]. Thus now a day's innovative

developers do not have to bother about huge investments on hardware setup and human resource before deploying any new service.

Cloud security can be defined as a composite notion, namely “the combination of confidentiality, the prevention of the unauthorized disclosure of information, integrity, the prevention of unauthorized amendment or deletion of information, and availability, the prevention of the unauthorized withholding of information”[18].

Security in Cloud Computing fall into two broad category- security issues faced by cloud computing providers and security issues faced by their customers. In most cases the provider must ensure that their information is secure and that their clients data and applications are protected while that customer must ensure that provider has taken the proper security measures to protect their information. Computer and network security is fundamentally about three goals or objectives- Confidentiality(c), Integrity(I) , Availability(A).

This paper is organized as follows: in the next section we describe the system architecture, then some applications of cloud computing. In the forth section we describe some security issues related to the cloud computing then after the next section enlist some related work. Then after we explained our proposed work and at last references after the conclusion of this paper.

2. SYSTEM ARCHITECTURE

To evaluate various policies for resource provisioning, load balancing, workload modeling and performance modeling, we need repeated testing under varying system and user configuration. It is always very tough to perform these testing in real practice and therefore there is need of simulation. Simulation is an act of imitating behavior of some process by means of something suitably analogous. Here we have used Cloudsim simulation toolkit [11] for simulating cloud environment.

In this work, our system comprises number of distributed serving nodes and users that can change their geographic location with time i. e. experiment will be performed on heterogeneous system. To present working in hierarchical manner, we choose three level architecture as shown in figure 1[13]. The first level is the Broker level, whenever a request arrives broker performs the task of datacenter selection on the basis of some parameter like least latency from user or minimum load on datacenter. Next is the datacenter level to decide which host will handle the request. At last host level where virtual

machines are created. Actual processing is done by virtual machine.

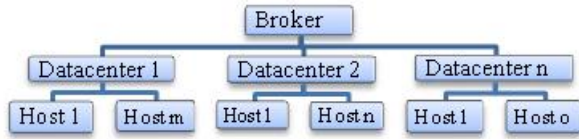


Figure 2.1 Three level system architecture

3. APPLICATION OF THE CLOUD COMPUTING

There is vast applications of Cloud computing. But recently there are two areas where cloud computing plays a major role which are forensics and Health sector. Cloud is becoming more and more important in the field of Health records as we know that the health sector has globalized so there is a mass of data is available and shared for which the health sector uses the cloud services [5].

In recent year, Personal health record (PHR) has emerged as a patient-centric model of health information exchange. A PHR service allows a patient to create, manage, and control her personal health data in one place through the web, which has made the storage, retrieval, and sharing of the medical information more efficient. Especially, each patient is promised the full control of her medical records and can share her health data with a wide range of users, including healthcare providers, family members or friends. Due to the high cost of building and maintaining specialized data centers, many PHR services are outsourced to or provided by third-party service providers, for example, Microsoft Health Vault. While it is exciting to have convenient PHR services for everyone, there are many security and privacy risks which could impede its wide adoption. The main concern is about whether the patients could actually control the sharing of their sensitive personal health information (PHI), especially when they are stored on a third-party server which people may not fully trust. [5]

The cloud computing is also used in the field of Forensics [4]. In the paper "Cloud Application Logging for Forensics" Raffael Marty stated that how cloud can be so useful in forensics. A cloud-based application stores logs on multiple servers and in multiple log files. The volatile nature of these resources causes log files to be available only for a certain period of time. Each layer in the cloud application stack generates logs, the network, the operating system, the applications, databases, network services, etc. Once logs are collected, they need to be kept around for a specific time either for regulatory reasons or to support forensic investigations.[4]

Security logging in cloud application is concerned with authentication and authorization, as well as forensics support. In addition to these three cases, security tools (e.g. intrusion detection or prevention system or antivirus tools) will log all kinds of other security-related issues, such as attempted attacks or the detection of a virus on a system. Cloud-applications should focus on the following use-

cases: Login / logout (local and remote), Password changes / authorization changes, Failed resource access (denied authorization), All activity executed by a privileged account, Privileged accounts, admins, or root users are the ones that have control of a system or application. They have privileges to change most of the parameters in the application. It therefore is crucial for security purposes to monitor very closely what these accounts are doing.

4. SECURITY ISSUES OF CLOUD COMPUTING

High security is one of the major obstacles for opening up the new era of long dreamed vision of computing as a utility[7]. As the sensitive applications and data are moved into the cloud data centers, and run on computing resources in the form of Virtual machine, but this unique feature, however poses many novel tangible (such as unauthorized access, Infrastructure failure, or unavailability) and intangible (such as confidence in the technologies capabilities and public access) security challenges.[18]. To protect personal and sensible data which is going to process at some data centers, the cloud users needs to verify,- the real exists of cloud computing environment in the world , the security of the information in cloud, and the trustworthiness of systems in cloud computing environment.CSA (Cloud Security and Alliance)[6], NIST(National Institute of Standard and technology)[2] these two organization has taken out some of the major security issues relevant to the cloud computing. As it is widely know factor that the major concern about cloud computing is the security. There are two aspects of cloud security that the security at the cloud consumer and security as the cloud service provider. Customers are both excited and nervous at the prospects of Cloud Computing. They are excited by the opportunities of Cloud Computing as it reduces capital costs, infrastructure management, and focus on core competencies. They are also excited as it is on-demand provisioning of computing and ability to align IT with business strategies. However, consumers are also very concerned about the risks of Cloud Computing if not properly secured and loss of direct control over systems for which they are nonetheless accountable.

So keeping both cloud consumer and cloud providers CSA developed "Security Guidance for Critical Areas in Cloud Computing" [2] initially released in April 2009, and revised in December 2009. These guidelines have become the industry standard catalogue of best practices to secure Cloud Computing. CSA defined "Top threats to Cloud Computing" are listed below:

- Abuse and nefarious use of cloud computing
- Insecure application programming interfaces
- Malicious insiders
- Shared technology vulnerabilities
- Data loss/leakage
- Account, service & traffic hijacking
- Unknown risk profile.

All of the above listed threats have serious impact to the cloud computing but Data loss or Leakage is what makes fearer to the consumers to put their data/information on the cloud. Data loss or Leakage can have a devastating impact

on a business. Beyond the damage to one's brand and reputation, a loss could significantly impact employee, customer morale and trust. There are many ways to compromise data as deletion or alternation of records without a backup of original content is an obvious example. The threat of data compromise increases in cloud, due to the number of transactions between risks and challenges which are either unique to cloud, or more dangerous because of the architectural or operational characteristics of cloud environment.

This data loss or compromises happened due to insufficient authentication, authorization and audit controls, inconsistent use of encryption and software keys. Data security involves encrypting the data as well as ensuring that appropriate policies are enforced for data sharing.[17] NIST also defined some key security and issues for the Cloud Computing security. Along with the issues it has also defined some guidelines to address those issues. Ensuring systems are secure and risk is managed is challenging in any environment and even more daunting with cloud computing. Audit mechanisms and tools should be in place to determine how data is stored, protected, and used, to validate services, and to verify policy enforcement. A risk management program should also be in place that is flexible enough to deal with the continuously evolving and shifting risk landscape. Various types of security and privacy laws and regulations exist within different countries at the national, state, and local levels, making compliance a potentially complicated issue for cloud computing.

One of the most common compliance issues facing an organization is data location. Use of an in-house computing center allows an organization to structure its computing environment and to know in detail where data is stored and what safeguards are used to protect the data. In contrast, a characteristic of many cloud computing services is that data is stored redundantly in multiple physical locations and detailed information about the location of an organization's data is unavailable or not disclosed to the service consumer. This situation makes it difficult to ascertain whether sufficient safeguards are in place and whether legal and regulatory compliance requirements are being met. Though Technical, physical and administrative safeguards, such as access controls, often apply.

One of the most common and very harmful security issues or we can say the most important threat to data stored is Insider access. Insider threats go beyond those posed by current or former employees to include contractors, organizational affiliates, and other parties that have received access to an organization's networks, systems, and data to carry out of facilitate operations. Incidents may involve various types of fraud, sabotage of information resources, and theft of sensitive information.[2]

For risk management organizations also needed to have knowledge of a cloud provider's security measures. By knowing security measure's one can take decisions to choose cloud applications to adopt if they fitted to their security measures. Data sensitivity and privacy of information is always an area of concern for any organization but now it is increasing rapidly. So before

moving to cloud services any organization has to take account all the security risk and issues. By providing authentication mechanism one can establish a confidence in user identities. [2]

Another security issue is Denial of service attack which also accounts a measure security concern in the cloud computing. This attack involves saturating the target with bogus or unwanted requests to prevent it from responding to legitimate requests in a timely manner. This attack mainly consumes large amount of resources which makes system slow down and make it vulnerable, to attack.

5. RELATED WORK

In this section of paper we discuss the related work which has been done in the field of Cloud Computing security. One of the most applicable methods for security is the Log Management. For the monitoring of the system activities the log is managed. Along with the log management the Cloud Computing system needs to contain Intrusion Detection System (IDS) for protecting the virtual machine against threats. IDSs are one of the most popular devices for protecting Cloud Computing systems from various types of attack. Because an IDS observes the traffic from each VM and generates alert logs, it can manage Cloud Computing globally. Jun-Ho Lee and his colleague had proposed a multi-level intrusion detection system and log management for implementing effective IDS in cloud computing system [9]. Cryptography is also one of the major technique in the field of security. Before cloud computing cryptography and at current time also cryptography is used to protect the information in the network. Now Cryptography can also be implemented in the cloud computing to protect the secret data [8]. There are many cryptography techniques have been invented some of which comes in the symmetric category and some of comes in the asymmetric category or we can say single key cryptography and public-private key cryptography. Both the techniques have their respective advantages and disadvantages, but practically both the techniques are used in a hybrid fashion. In cloud computing cryptography can provide the privacy and integrity of the data storage so that cloud computing can be adopted at areas where information security matters a lot. The main concern of data storage is the protection of information from unauthorized access. In this way cryptography can play a major role to protect from unauthorized access of information and also about its integrity. The main issue in the secure storage approach is the management of encryption key. In fact once the data is encrypted the key became the true bits to protect. If he keys are stored on un-trusted environment along with the data then, an attacker could have access to that key so in case will be able to access the encrypted data. So to solve that problem keys are stored in keystore that can be implemented either in portable device by the use who can plug it anywhere or in a specialized server which sits somewhere in trusted environment. New technique is derived which greatly resolve the problem of sharing of shared secret key, which is searchable encryption. Searchable encryption is a broad concept that deals with searches in encrypted data. The goal is to outsource

encrypted data and can be able to conditionally retrieve or query without having to decrypt all the data [8].

With the cryptography, log management and IDS is also equally important part of security and this provide so much of security to the cloud computing. Cryptography provides the secrecy and integrity of data but we also need to tracking of our information/data and this facility is provided by the log management. The need for tacking and monitoring is pervasive throughout many aspects of an organization. Logs contain important information on performance, security and user activities, and provide critical information about what the system is doing. With migration to the cloud, logging and monitoring became even more important since you are giving up some control to the cloud service provider [19]. Cloud based applications stores logs on multiple log files. Once logs are collected they need to be kept around for a specific time either for regulatory reasons or to support forensic investigations. Events should not be captured at a too abstract level as that will result in an audit trail which will be too general for forensic use. On the other hand, the events also cannot be captured at a too detailed level as that will result in an explosion in the size of the log files generated.

Yu Shyang Tan and his colleague published a paper regarding tracking of data leaving the cloud. Is this paper he had mentioned many things like requirement of data tracking framework. So why be need data tracking framework there are some points which are: tracking of data, sending an event log back to the server, storing and processing the event log.

6. PROPOSED WORK

Here we drew the limitations and drawback of previously used data tracking system, and also proposed a new optimized tracking system. In the previous method of data tracking proposed by Yu Shyang Tan with his team members at HP Labs, Singapore and HP Software, CA, USA, actions or events, such as access and modification, to data residing outside the cloud can be tracked, logged and sent back to the main server. Following which they show how the captured information can be combined with the event logs of the data. The method having two components which are 1) Data tracker, and 2) Event Detection. As a data tracker mechanism they make use of archival techniques to archive the data into a self-executing container to form a medium that is able to follow the data tracking purposes and to protect the data from direct accesses. This is done at the point when the data is about to be moved from the cloud.

When a user attempts to copy out or remove one or more data files out of the cloud boundaries, from either the VMs or the storage medium in the cloud a “check-out” via a provided web interface is enforced. The selected data files are encapsulated together with a viewer program, into a self-executing container. The self executing container will execute the viewer program that resides within the container. Once the various data files and the viewer program are archived into the container. The container is encrypted to prevent users from having direct access to data

files since only the viewer will have the key, which is pre-programmed into it at the point of creating the container, to decrypt the encryption. The viewer program is a small program which serves two purposes. First, it acts as an interface through which users can interact with the data and perform actions such as view, modify or delete, on the data. Activities will be monitored by the viewer program. That way, the second purpose of the viewer program can be done, which is to captured and log down the events being captured. An overview can be depicted as shown in figure 6.1.

From the figure 6.1 we can understand that how the viewer program communicates with the cloud. This viewer program is the entry of the stored data, and this also serves its second part that log generation and send back to the server.

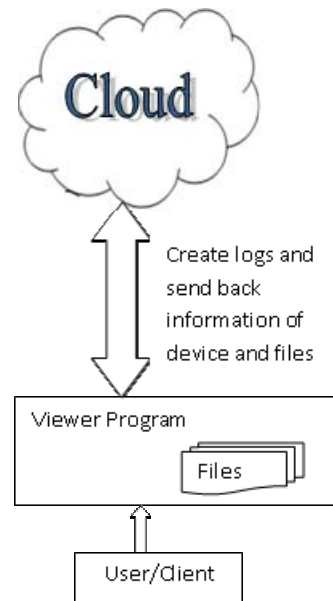


Figure 6.1 Data tracker module

To provide the security to the logs before sending back to the server this log file is encrypted with either symmetric key cryptography or private key cryptography such as AES and RSA. The log generated by the viewer program contains all the necessary information so that it will be easy to track that who access the data, when access the data, where access the data, what has been done to that particular data files. So we can see this all by the following diagram that is kept by the server when this log file is send back to the server. This all information regarding to the user who accessed the data files is stored in the database.

Along with this, things i.e data tracker and log file generation this does not gives us facility that we can recover our data to its original position and this was the major disadvantage to that data tracking mechanism. So to remove that disadvantage we have proposed a new technique that can recover our data to its original position. So that if there is some unwanted modification by some unauthorized access of data and information, it will be tracked and we will be in a position to recover data. So that we can say this technique prevent the data loss. In this mechanism we also captured one more information along with all user information is the, “is data files is modified”.

And also here we provide a copy of data files to the user not the actual data. So the owner of data can make sure that his/her information or stored data did not compromise.

So as a whole in this mechanism first thing is data is kept inside a data tracker which contains a viewer program, a user who gets access to the data files got logged and that log file is encrypted before sending back to the server, just to provide the integrity. of the log files. When this log file is send back to the server it is then decrypted accordingly as it was encrypted. Then after this log information is stored to the database as to prevent log information. As long as this log information is stored into the database we also check that, is files has modified or not? if yes then we also check that, are the modifications correct or not? After checking this if the modification into the data are correct

then that modified copy is replaced with the old copy. So now that newly replaced file is available to the other users. Plus one more thing if the modifications are not correct and can say that if then access was not authorized then the data owner can take serious action against the attacker. In this method the log files are saved just for a certain amount of time. After that this log files is replaced by some other log files, because if we keep all the log files are saved then it may create the huge amount of log information.

So as to solve this problem the log files are immediately send to the server and that information is saved into the database. This prevents the loss of information and also reduces the unnecessary space which was used to keep log information.

username	ip	hostname	macaddress	platform	datetime	filename	modified
lucky	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:41:20 IST 2014	build.xml	yes
lokendra	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:42:13 IST 2014	app.txt	yes
lokendra	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:51:28 IST 2014	app.txt	no
admin	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:54:39 IST 2014	build.xml	yes
admin	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:56:20 IST 2014	app.txt	yes
admin	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 13:57:23 IST 2014	app.txt	no
lokendra	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 14:01:22 IST 2014	build.xml	yes
lokendra	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 14:32:31 IST 2014	build.xml	yes
lokendra	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 14:36:05 IST 2014	app.txt	yes
rahul	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 14:54:59 IST 2014	build.xml	yes
amit	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 14:56:59 IST 2014	build.xml	yes
admin	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 15:06:23 IST 2014	app.txt	yes
admin	172.16.11.217	Lokendra-PC	24-B6-FD-54-4F-3F	Windows 7	Fri Jan 24 15:09:09 IST 2014	build.xml	yes

Figure 6.2 Log information saved in the database

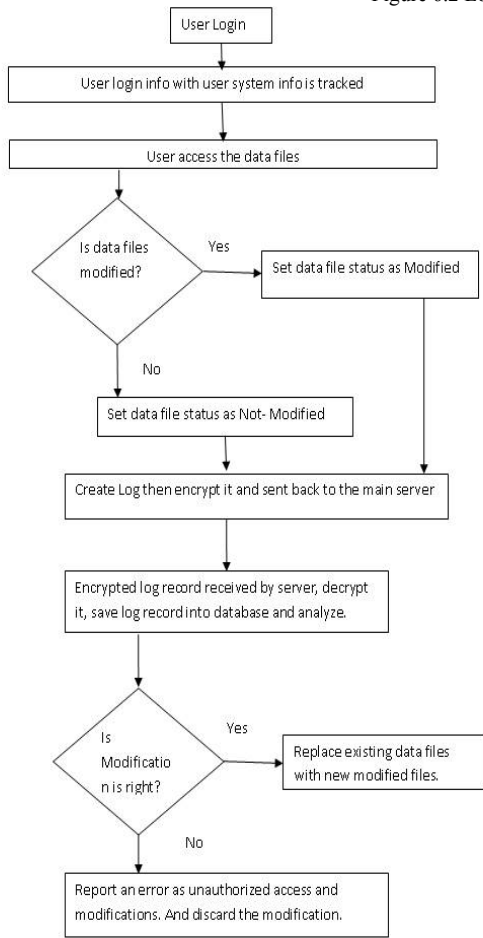


Figure 6.3 Flow chart of proposed method.

7. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an Optimized method for tracking data leaving the cloud environment. Using this method we can assure the user or organization who wanted to use the cloud services to store data but because of lack of security and lack of data control they generally not using technology. Results show that new optimized method gave more control over the data and more security to the data. This method provides a facility to the data owner that in any case if there is some unwanted modification to the data, data will be recovered and the information and data is kept inside a data tracker in encrypted form, which gave it some more security. Each and every time data is accessed a log is created whether the data files is modified or not so which gave a track record that who access what and where. Although proposed algorithm provides good results for security and tractability, but there is some scope to improve its efficiency and security by applying some other security methods. The main concern is the viewer program or can say the data tracker needs to be more secure so that in any case this will never be compromised. So in future this method would we more improved.

REFERENCES

[1] Yu Shyang Tan, Ryan K L Ko, Peter Jagadpramana, Chun Hui Suen, Markus Kirchberg, Teck Hooi Lim, Bu Sung Lee, Anurag Singla, Ken Mermoud, Doron Keller, Ha Duc, "Tracking of Data Leaving the Cloud", "2012 IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications". 978-0-7695-4745-9/12, DOI 10.1109/TrustCom.2012.282

[2] Guidelines on Security and Privacy in Public Cloud Computing, SSpecial Publication 800-144 Wayne Jansen, Timothy Grance, Computer Security Division, Information Technology Laboratory

National Institute of Standards and Technology, Gaithersburg, MD 20899-8930

- [3] Shilpashree Srinivasamurthy and David Q. Liu, "Survey on Cloud Computing Security", Department of Computer Science, Indiana University – Purdue University Fort Wayne, Fort Wayne, IN 46805.
- [4] Raffael Marty Loggly, Inc. "Cloud Application Logging for Forensics" SAC'11 March 21-25, 2011, TaiChung, Taiwan, ACM 978-1-4503-0113-8/11/03
- [5] Ming Li, Shucheng Yu, Yao Zheng, Kui Ren, and Wenjing Lou, "Scalable and Secure Sharing of Personal Health Records in Cloud Computing Using Attribute-Based Encryption" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 1, JANUARY 2013, 1045-9219/13
- [6] Cloud Security Alliance, "Top Threats to Cloud Computing V1.0", March 2010. <http://www.cloudsecurityalliance.org/topthreats>
- [7] Dawei Suna,, Guiran Chang, Lina Sun and Xingwei Wang, "Surveying and Analyzing Security, Privacy and Trust Issues in Cloud Computing Environments", ELSEVIER.
- [8] Isaac Agudo , David Nuñez , Gabriele Giammatteo , Panagiotis Rizomiliotis , Costas Lambrinouidakis, " Cryptography goes to the Cloud".
- [9] Jun-Ho Lee, Min-Woo Park, Jung-Ho Eom, and Tai-Myoung Chung, "Multi-level Intrusion Detection System and Log Management in Cloud Computing"
- [10] Mirko Montanari, Jun Ho Huh, Derek Dagit, Rakesh B. Bobba, Roy H. Campbell, "Evidence of Log Integrity in Policy-based Security Monitoring".
- [11] Rodrigo N. Calheiros, Rajiv Ranjan, César A. F. De Rose, and Rajkumar Buyya, "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms", ACM journal of Software-Practice & Experience, vol.41 issue 1, pp. 23-50, Jan. 2011.
- [12] Rodrigo N. Calheiros, Rajiv Ranjan, César A. F. De Rose, and Rajkumar Buyya, "CloudSim: A Novel Framework for Modeling and Simulation of Cloud Computing Infrastructures and Services",
- [13] S. Wang, K. Yan, W. Liao and S. Wang, "Towards a Load Balancing in a three-level cloud computing network", 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, pp. 108-113, 2010.
- [14] Jie Tao, Holger Marten, David Kramer and Wolfgang Karl, "An Intuitive Framework for Accessing Computing Clouds", ELSEVIER International Conference on Computational Science (ICCS), pp. 2049–2057, April 2011.
- [15] Tarun Goyal, Ajit Singh, Aakansha Agrawal, "Cloudsim: simulator for cloud computing infrastructure and modeling", ELSEVIER International Conference on modeling, optimization and computing (ICMOC), vol. 38, pp. 3566-3572, 2012.
- [16] Suraj Pandey, William Voorsluys, Sheng Niu, Ahsan Khandoker, Rajkumar Buyya, "An Autonomic Cloud Environment for Hosting ECG Data Analysis Services", The University of Melbourne, Australia.
- [17] Kevin Hamlen, Murat Kantarcioglu, Latifur Khan, Bhavani Thuraisingham "Security issues for Cloud Computing", The University of Texas at Dallas, USA.
- [18] Scott Paquette , Paul T. Jaeger, Susan C. Wilson "Identifying the security risks associated with governmental use of cloud computing" ELSEVIER International conference, 13 April 2010.
- [19] Logging and alerting for the cloud, enstratus, Inc
- [20] .