iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** XI **Month of publication:** November 2019

DOI: http://doi.org/10.22214/ijraset.2019.11076

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XI, Nov 2019- Available at www.ijraset.com

Data Mining Techniques with Cloud Security

Mrs. S. Revathi[1],
Assistant Professor (SG), Department of Computer Science, Dr.N.G.P Arts and Science College, Coimbatore

Abstract: Cloud computing has now extended as a new paradigm for delivering services over the Internet. It is attractive to business owners, since it reduces the upfront investments. However, despite the fact that cloud computing offers potential opportunities to the IT industry, it has many issues and challenges still to be addressed. Eventually the business owners could not seek the physical location of their application or data, which reduces the confidentiality on cloud computing. Audit logs are used as one of the possible ways to secure the data and it allows forensic analysis when any security incidents that occurs. This paper present the data mining techniques used in audit logs.
Keywords: Cloud Computing, Data Mining, Audit Logs

I. INTRODUCTION

The cloud computing concept is evolved as an Internet based pre-configured computing environment, which can be accessed on-demand and ubiquitous model.

National Institute of Standards and Technology (NIST) defined the cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources such as networks, servers, storage, applications and services that can be rapidly provisioned and released with minimal management effort or service provider interaction"[1].

It is mainly attract the software business vendors through its rapid provisioning and releasing capabilities. Further, it reduces the upfront infrastructure cost to the companies. Therefore, the cloud computing is relatively economy and convenient than the conventional method[2].

The cloud computing service model is generally offered in three different types known as software as a service (SaaS), platform as a service (PaaS) and infrastructure as a service (IaaS). Security is a common challenge among those services in cloud computing. Security audit is a straight forward mechanism to investigate the happenings and prevent through right approach. Especially in the information technology industry, the audits are typically executed in two different patterns known as internal and external. The internal audit refers to work done by an organization's own employees, whereas the external audit refers to the third party involvement in the auditing activities.

Existing security policies are not capable in the cloud computing context. In order to manage the confidentiality and challenges, cloud advocates such as Cloud Security Alliance (CSA)[3] are urging standardization of cloud confidentiality, integrity and availability. The remaining part of this paper presents the cloud computing, cloud security auditing, audit logs, data mining process and mining audit logs.

II. CLOUD COMPUTING

Cloud computing is a nascent and rapidly evolving model, with new aspects and capabilities being announced regularly. Cloud computing is the result of the evolution and adoption of existing technologies and paradigms. The goal of cloud computing is to allow users to take benefits from all of these technologies, without the need for deep knowledge about or expertise with each one of them[4].

The cloud enhances collaboration, agility, scalability, availability, ability to adapt to fluctuations according to demand, accelerate development work, and provides potential for cost reduction through optimized and efficient computing [5,6]. Cloud computing is a combination of different technologies such as virtualization, Web 2.0, Service oriented architecture and many more. Cloud computing has three distinct service models and three delivery models.

The cloud services are offered in three different models and this section describes its characteristics. Subsequent Figure-1 illustrates the cloud service architecture.

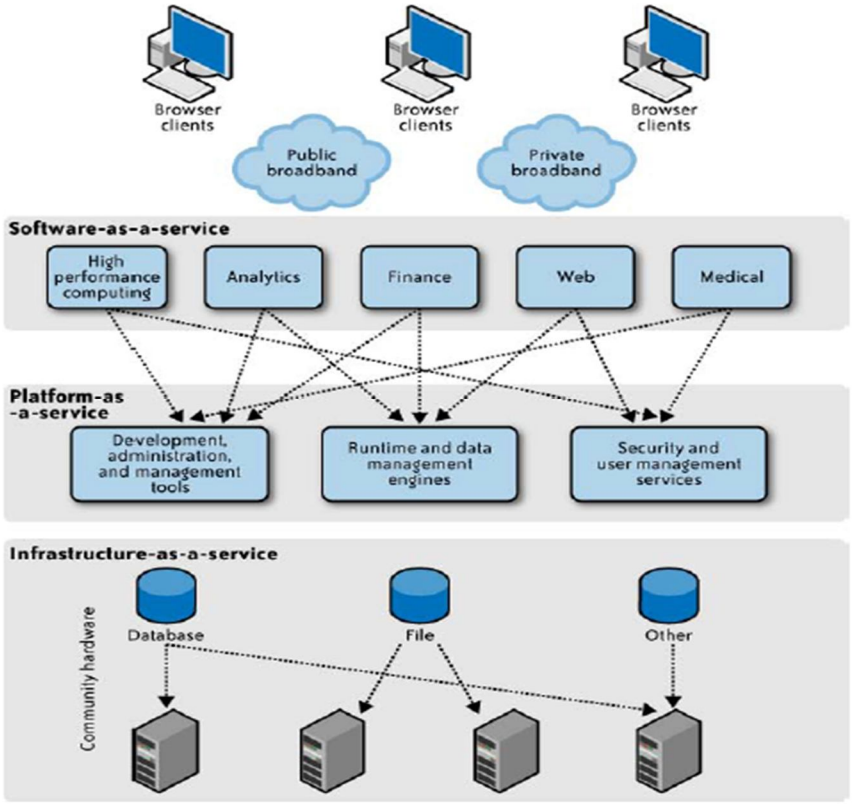


Figure-1: Cloud Service Architecture

1) Infrastructure as a Service Model, service provider provides virtual and physical hardware as a service and entire infrastructure is delivered over the internet. In this model client has more security control. Provider provides networking, virtualization, servers and storage [5]. The characteristics of Iaas are:

a) Utility computing service and billing model.
b) Automation of administrative tasks.
c) Dynamic scaling.
d) Desktop virtualization.
e) Policy-based services.
f) Internet connection

2) Platform as a Service Model provides for development and deployment software applications by supporting entire application cycle. Cloud provider is responsible for security and monitoring. Provider provides runtime, middleware, OS, networking, servers, storage and virtualization. Developer takes several benefits from Paas. OS features could be easily changed with Paas[7,8]. Geographically distributed development tea can obtain service from diverse source and work together on software development projects.

3) Software as a Service Model, consumer use hosted application through a web browser. In the SaaS model, Security, management and control are services provider's responsibility because the customer has minimal control or extensibility. By contrast, the Paas model offers greater extensibility and greater customer control [9, 10]. Largely because of the relatively low degree of abstraction, Iaas offers greater tenant or customer control over security than does PaaS or SaaS [11]. Characteristic of SaaS are:

a) Computerized billing
b) Invoicing
c) Human Resource management
d) Collaboration
e) Document Management
f) Service Disk Management

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 7 Issue XI, Nov 2019- Available at www.ijraset.com

As like service model, the cloud computing has three deployment models such as private cloud, public cloud and hybrid cloud. The public and private cloud concepts are important for organizations which share its resources over Internet to customers for a fee. The end users who use the services offered via cloud computing may not have knowledge about the deployment backbone.

i) *Private Clouds:* are known as internal clouds which offering the cloud computing on private networks. Organizations must buy, build and manage the setup by own. The organizations have more control over physical and logical security [16, 17].
ii) *Public Clouds:* are known as external clouds which is a widely adopted system of cloud computing. User organizations can avail this feature at lower upfront capital costs. Security management and day-to-day operations are relegated to the third party vendors. Hence, the customer has less degree of control over physical and logical security aspects of a private cloud.
iii) *Hybrid Clouds:* consisting of multiple internal and external providers in a deployment possible for organizations.

A. Challenges in Cloud Computing

The cloud computing is generally has its own security challenges. Especially in the cloud infrastructure, the cloud service should constantly negotiate with three entities such as service organization, cloud service provider (CSP) and end users. Even though CSP has given enough attention on securing data, security threats are unavoidable. Since, it can be accessed by the client from anywhere through Internet. Cloud user organizations have no idea on the physical location of their servers or application, which leads to the transparency issues. In the similar-way the encryption policy maintained by the CSP is not foolproof. If the cloud breached, the information in the server would be instantly available for hackers. There are different methods exists to ensure the cloud security such as infrastructure security, data security, storage security, identity & access management security and audit & compliance.

III. CLOUD AUDIT LOGS

Audit log is a document which records the events happening in the system. The audit log mainly includes the source, destination IP address, timestamp, users and login instructions[15]. The logs entries are important for security and process improvement reasons which are broadly classified into four groups[16].

1) *Accountability:* Log data can identify what accounts are associated with certain events. This information then can be used to highlight where training and/or disciplinary actions are needed.
2) *Reconstruction:* Log data can be reviewed chronologically to determine what was happening both before and during an event. For this to happen, the accuracy and coordination of system clocks are critical. To accurately trace activity, clocks need to be regularly synchronized to a central source to ensure that the date/time stamps are in synch.
3) *Intrusion Detection:* Unusual or unauthorized events can be detected through the review of log data, assuming that the correct data is being logged and reviewed. The definition of what constitutes unusual activity varies, but can include failed login attempts, login attempts outside of designated schedules, locked accounts, port sweeps, network activity levels, memory utilization, key file/data access, etc.
4) *Problem Detection:* In the same way that log data can be used to identify security events, it can be used to identify problems that are need to be addressed (i.e. investigating causal factors of failed jobs, resource utilization and trending).

The data logged in a system must be retained for long enough to answer many questions and the log information are useless when no reviews happened.
Maintaining the log information consume storage space, cost and it should be maintained at optimum level.
A cloud audit log entry is essentially a record of an event where cloud data was accessed or modified at any given time. All of those instances are tracked and store the cloud audit logs. If at any time in the future an organization needs to prove that a specific event occurred in the cloud, it can present the audit logs as evidence. Furthermore, regulations require that organizations properly maintain and store their audit logs so that they may be easily accessible in the future, for instance during a forensic investigation or during electronic discovery.
Essentially, audit logs benefit both the cloud consumer and cloud provider in that transparency in cloud computing is properly being recorded and maintained. Proper cloud log management is essential for any organization benefiting from cloud computing services, regardless of its size and scope.
As companies begin to migrate their data to the cloud, they may overlook the importance of archiving cloud audit logs on a regular basis and compliance regulations associated with them.

IV. DATA MINING METHODS

The increasing ability to generate massive amount of data brings potentials to discover and utilize the knowledge extracted from data. Data mining has been identified as an effective tool to analyze data[14]. It is a type of database analysis that attempts to discover useful patterns or relationships in a group of data. The analysis uses advanced statistical methods, such as cluster analysis, and sometimes employs artificial intelligence or neural network techniques. A major goal of data mining is to discover previously unknown relationships among the data, especially when the data come from different datasets. Various data mining methods are available to extract useful in information such as association rule mining, classification, clustering, decision tree, factor analysis, genetic algorithms, neural networks, outlier detection, regression analysis, sequence mining, support vector machines, text mining and agent mining.

1) *Association Rule Mining:* is a method for discovering interesting relations between variables in large databases based on the rules specified. This data mining method is widely applied to extract the behavioural pattern of data.
2) *Classification Mining:* is a process of identifying the new data category based on the training set. This data mining method is very useful in pattern recognition.
3) *Cluster Analysis:* is a method of categorizing data into subsets based on the specification. Majority of the cases, the cluster notation cannot be precisely defined.
4) *Decision Tree:* is a type of decision support tool that uses a tree-like graph to identify the right decision among the possible alternatives. This mining method is extensively used in the operation research.
5) *Factor Analysis:* method is used to measure the variability among observed and correlated variables.
6) *Genetic Algorithm:* is a method for solving both constrained and unconstrained optimization problem that is based on the natural selection.
7) *Neural Network:* Is a network inspired by biological neural networks which are used to estimate unknown pattern.
8) *Outlier Detection:* is also known as anomaly detection which tries to identify the patterns of data that are deviated from the remaining data. This method is frequently used to identify bank frauds, network intrusions and fault detection.
9) *Regression Analysis:* Is a method of estimating the relationship among dependent and independent variables, which is mainly used for designing a model or framework.
10) *Sequence Mining:* is a statistical based pattern mining method which helps to extract the frequently occurring patterns.
11) *Support Vector Machines:* is a supervised learning method and it uses classification and regression analysis. It is primarily used discriminate the data.
12) *Text Mining:* is a method of discovering high quality information from the given text. It is also related to statistical pattern mining technique and widely applied in the information retrieval and natural language processing areas.

Data mining generally helps to discover the pattern and useful information in the given dataset. The main intention of this paper is to exhibit the possible data mining methods for cloud audit logs and the subsequent section discuss in detail.

V. MINING CLOUD AUDIT LOGS

Cloud systems often continuously record the logs at different levels, which provide useful information to maintain the complex computing infrastructure. Many cloud app security keeps detailed log about activities affecting the management console, threat protection, data loss prevention. Further, the log reports given by the systems help to mitigate the threats and optimizing the system settings.

Audit logging is critical for user organizations and required for many regulatory bodies. Enterprises especially in regulated industries that need to show accurate logs of data and application access hesitate to use SaaS for this very reason. Most SaaS applications today are like a black box for customer organizations which has not provided visibility into what their users are doing in the cloud. Meanwhile for some enterprises this is a definite show stopper for compliance reasons or in some cases internal security policies. Eliminate these concerns by tracking and reporting access events at a page level; from the cloud application back to the enterprise. Cloud Security Service can provide deep page level audit tracking and can offer to customers as value added compliance reports.

Log mining is technically a kind of analysis which helps to monitor the network, system health, application and users. While making log analysis, certain constrains are exists for instance log data is generally too large, it contains missing data, diverse records, false alarms and duplicate data. Considering that mining in the cloud is something new, still there is no large number of solutions that are fully completed and available to users. Google BigQuery service, Amazon Elastic MapReduce and MS-SQL Server Data Mining are readily available in the market. Google cloud audit logging is primarily considered in this paper. Google cloud logging consists of

two log streams such as admin activity and data access. The admin activity logs contain an entry for every API call and administrative actions that modifies the configuration or metadata of a service. This log is visible to all members in involved in this project. Data access logs on the other hand contain entry for API call and administrative actions that create, modify or read user provided data managed by a service. These logs are visible only to the project owners and users. Individual audit log entries are kept for a specified length of time and are then deleted. The Stack driver Logging Quota Policy explains how long log entries are retained. Audit logs record the email address of the user who performs logged actions in the Authentication Info field of Audit Log objects. Audit logs can be viewed through LogViewer. Similarly, Audit logs can be exported as JSON object. The exported file can be applied for data mining to extract new insights from data.

Data mining technique anomaly detection is applied in the audit logs to detect the unusual or suspicious cases from the collection. In the application level, it helps to monitor the fraud detection, healthcare and expense report. In the network level, it helps to predict the intrusion detection and denial of service attacks. Audit logs can be further mined through classification, clustering techniques, genetic algorithm, neural networks, support vector machines and text mining.

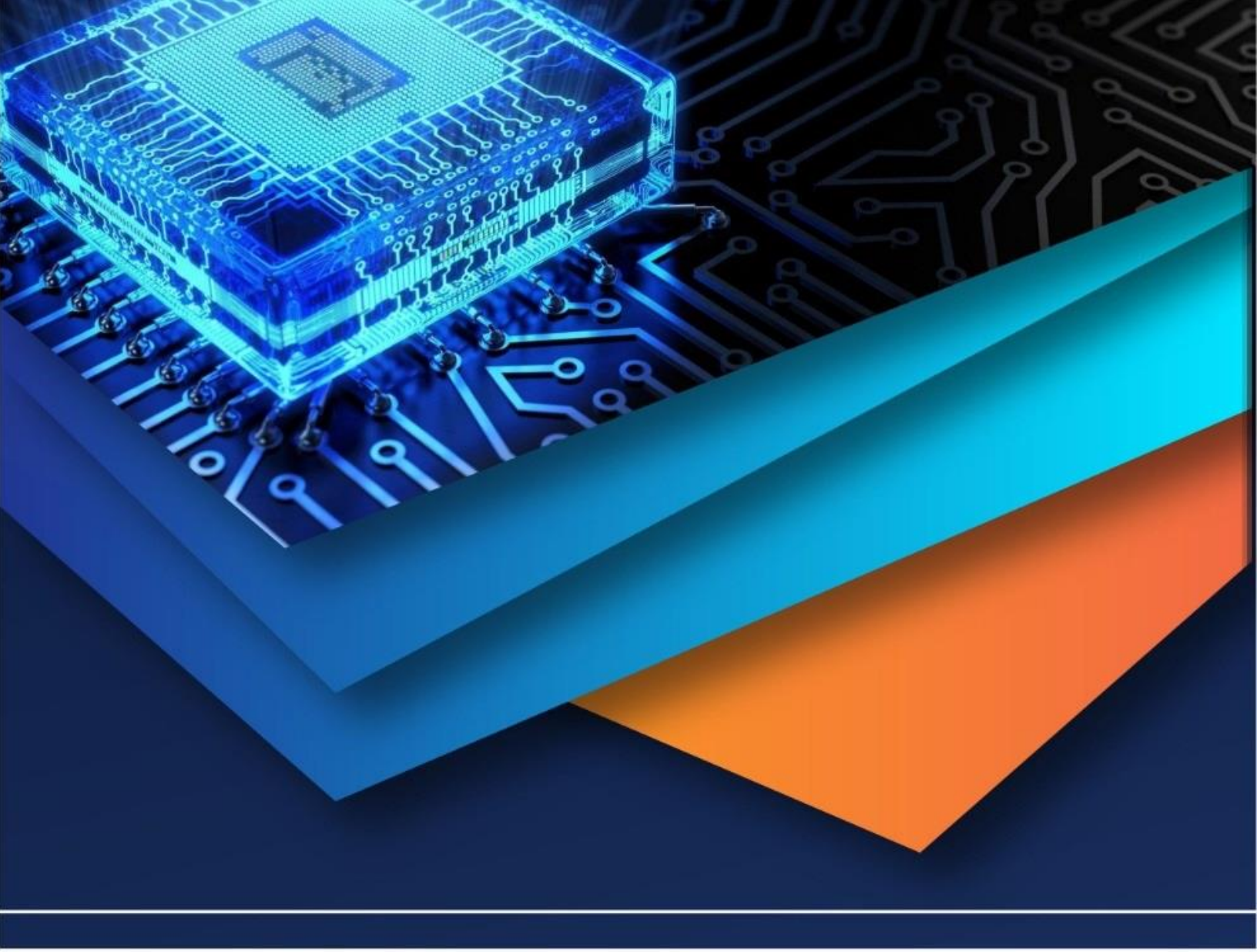
Classification mining can be applied to audit logs for extracting the new pattern from the patterns which are trained earlier. The classification may be based on the payload size, metadata and frequency of log entry and so on. After classification clustering can be applied to discriminate the data into subsets. In the similar fashion, support vector machines also used to discriminate the data in linear and non-linear perspective. Genetic algorithm and neural networks kind of data mining techniques also used in the audit logs to design a security policy and access parameters. The log and text payload fields in audit logs are string data type. Text mining can be applied to ascertain the textual information from the data through statistical pattern mining method.

VI. CONCLUSION

The cloud computing is naturally the complex infrastructure which makes the user organizations to avail the resources in convenient way. In this paper, general characteristics of cloud computing, data mining techniques and audit logs are discussed. Google cloud platform based audit logs are primarily consider for reviewing the data mining techniques applicable to mine the audit logs. The final outcome of this review states that audit logs are essential to monitor the administrative and user activities. Further, the data mining technique are used to prevent losses and helps to create the security, access policy.

REFERENCES

[1] Nist.gov: NIST SP800-53: security and privacy controls for federal information systems and organizations.
[2] Zhao G, Liu J, Tang Y, Sun W, Zhang F, Ye X, Tang N (2009) Cloud Computing: A Statistics Aspect of Users. In: First International Conference on Cloud Computing (CloudCom), Beijing, China. Springer Berlin, Heidelberg,pp 347–358
[3] http://www.cloudsecurityalliance.org
[4] Zhang S, Zhang S, Chen X, Huo X (2010) Cloud Computing Research and Development Trend. In: Second International Conference on Future Networks (ICFN'10), Sanya, Hainan, China. IEEE Computer Society, Washington, DC, USA, pp 93–97
[5] Cloud Security Alliance (2011) Security guidance for critical areas of focus inCloud Computing V3.0
[6] Khalid A (2010) Cloud Computing: applying issues in Small Business. In:International Conference on Signal Acquisition and Processing (ICSAP'10),pp 278–281
[7] Kitchenham B (2004) Procedures for performing systematic review, software engineering group, University of Keele, United Kingdom and Empirical Software Engineering, Australia.
[8] Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering. University of Keele, Department of Computer Science, UK
[9] Zissis, Dimitrios, and Dimitrios Lekkas. "Addressing cloud computing security issues." Future Generation Computer Systems 28.3 (2012): 583-592.
[10] Bhadauria, Rohit, and Sugata Sanyal. "Survey on Security Issues in Cloud Computing and Associated Mitigation Techniques." International Journal of computer applications, (2012).
[11] Harnik, Danny, et al. "Secure access mechanism for cloud storage." Scalable Computing: Practice and Experience (2011).
[12] Pappas, Vasilis, et al. "CloudFence: Data Flow Tracking as a Cloud Service". Research in Attacks, Intrusions, and Defenses. Springer Berlin Heidelberg, 2013. 411-431.
[13] Seccombe, A., et al. "Security guidance for critical areas of focus in cloud computing". Cloud Security Alliance (2009).
[14] Petre, Ruxandra-Stefania. "Data mining in cloud computing." Database Systems Journal 3.3 (2012): 67-71.
[15] Xie, Rui, Rose Gamble, and Norman Ahmed, "Diagnosing Vulnerability Patterns in Cloud Audit Logs", High Performance Cloud Auditing and Applications, Springer New York, 2014, 119-146.
[16] http://www.datamation.com/columns/article. php/3578916/The-Importance-of-Audit-Logs.htm



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY