Journal of Theoretical and Applied Information Technology
20th January 2014. Vol. 59 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EMPLOYING DATA MINING ON HIGHLY SECURED PRIVATE CLOUDS FOR IMPLEMENTING A SECURITY-AS-A-SERVICE FRAMEWORK

[1]Prof. G. Thippa Reddy, [2]Prof. K. Sudheer , [3]Prof. K Rajesh, [4]Prof. K. Lakshmanna
[1]Assistant Professor, SITE, VIT University, Vellore
[2]Assistant Professor, SITE, VIT University, Vellore
[3]Assistant Professor, SITE, VIT University, Vellore
[4]Assistant Professor, SITE, VIT University, Vellore

ABSTRACT

Cloud computing is rapidly gaining popularity. However, like any new system, cloud computing is facing some significant challenges. The most significant challenge faced by cloud adopters is related to legal compliance, security controls, privacy and trust. This paper proposes a novel security management framework by employing data mining to detect, contain and prevent attacks on cloud computing systems. Given that the security frameworks and related controls on cloud computing are still evolving, this research may prove to be a useful piece of contribution in enhancing the theoretical foundation established by existing studies.

Keywords: *Cloud Computing, Data Mining, Security-as-a-service, OPNET, Positivism and the Interpretivism*

1. BACKGROUND AND CONTEXT

A new concept of deploying IT services has emerged in past few years, which is based on the success stories of GRID computing [12]. The concept, popularly known as cloud computing, makes use of virtualisation like that in GRID computing, but has a better implementation of web services architecture. Technically speaking, cloud computing makes use of the same concept of distributed computing, as that in GRID computing albeit with employment of modern web 2.0 based services that were absent during the GRID computing days. In its current form, cloud computing offers many advantages to small and medium scale businesses. [23] and [7] presented the following advantages of cloud computing for small and medium enterprises

(a) The IT function can ensure rapid deployment of enterprise-wide applications without making investments in underlying hardware, software platforms and networking.
(b) The cloud service providers are capable of ensuring excellent service levels comprising high performance, high uptime, unlimited flexibility and scalability, and error-free operations.
(c) The cloud service providers provide platform and application services in such a way that they are independent of underlying hardware platforms. Hence, the system designs and operating processes can be customised as per client needs.
(d) All support services and development services are centralised by the cloud service providers. The end-customers need not hire IT developers and systems specialists.
(e) The end-customers can demand capacity as needed (elasticity) without making any capital investments.
(f) All overheads related to systems operations and maintenance are owned by the cloud service providers.
(g) The end-customers need not worry about obsolescence and continuous upgrading given that they are owned by the cloud service providers.
(h) The recurring costs are much lesser because the cloud providers charge rentals based on fixed resource or pay-per-use billing.
(i) The development platforms prevailing on the clouds are different and much simpler than the traditional development platforms.
(j) End-customers can enjoy latest software versions without paying for upgrading them.
(k) Some of the complex software systems, like SAP, Sales Force, and IBM Cognos, were out of reach of small and medium scale enterprises due to very high

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

licensing and implementation costs. Through cloud computing, they are easily available to them as well as to micro-scale enterprises.

The advantages of cloud computing listed above appear to be solutions to majority of the traditional IT related problems faced by enterprises in the past. Hence, cloud computing is rapidly gaining popularity. However, like any new system, cloud computing is facing some significant challenges. The most significant challenge faced by cloud adopters is related to legal compliance, security controls, privacy and trust. This has been highlighted by a number of papers, like [26], [17], [4], [27], and [3]. Some of the significant security, privacy, trust, and compliance related challenges discussed by these scholars are listed below:

(a) Cloud computing employs virtualisation at the core for pooling of system and software resources, and web 2.0 based services architecture for authentication, authorisation and provisioning services to the clients. The systems and applications penetration attacks in virtualisation and web 2.0-based architecture are new to the security administrators. The technical and operating level strategies for detecting, de-fusing and preventing such attacks in virtualisation and web 2.0-based architecture are not standardised yet. NIST has made their first ratified security standard for cloud computing public[15], but the standard is not adequate yet to answer all the cloud computing security-related concerns.

(b) The existing network and systems level intrusion detection and prevention tools are not effective in detecting exploits in a virtualisation environment launched from one of the virtual machines within the arrays. Given that the cloud services are open for everyone, an attacker can gain access to a virtual machine through a valid subscription and use it as a launch pad for diffusing malicious codes in other virtual machines in the array, and beyond. The interconnectivity among virtual machines is established through virtual networking that cannot be monitored by traditional network monitoring tools.

(c) There is no physical segregation of IT resources between any two tenants on the cloud. All segregations are deployed employing virtual boundaries, which makes access control settings quite complex. The concept of virtual segregation and security controls to protect virtual boundaries are not yet standardised adequately.

(d) Most of the clouds are created using internally developed models by the service providers. A draft of a standardised cloud computing model is released by NIST in 2011, described by Kaufmann et al.[16]. The proposed standard is de-facto and hence may be a good step forward to standardise cloud adoption by businesses and government organisations. However, it is very early to analyse the effectiveness of this standard from security perspective given that it is not yet ratified.

(e) Tenants on the cloud cannot trust the controls accessible and manageable by them in its current state of affairs.

(f) The existing standards for auditing and forensics are not suitable for cloud computing environment.

(g) Data proliferation is a currently an unmitigated risk given that the security levels at the virtual boundaries are not yet standardised.

(h) Cloud computing employs a large number of servers, storage and network devices distributed in the form of arrays. The virtualisation technology ensures that every piece of data is broken into fragments and distributed among the virtual servers within a virtualised cluster. Hence, it is almost impossible to locate data on the cloud physically. This may cause a boundary breach as data can be stored on servers outside national boundaries, which is illegal in many countries.

(i) Clouds are formed through collaboration of multiple service providers in the form of software as a service, infrastructure as a service and platform as a service, as described in [16]. There is a threat of inadequate or absence of trust relationship among the service providers having different roles on the cloud.

(j) There may be a threat of vendor lock-in on the cloud.

Security on the clouds should be distributed across all components possessing a multi-level

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

hierarchical architecture[10]. The controls cannot be placed at a distance away from the cloud arrays. They need to be embedded deeply into the arrays in the form of security applications. The NIST SP 800-144 standard proposed security controls in cloud apps design, cloud apps deployment, cloud apps distribution, access control through apps, and appropriate controls at platforms level, resource pooling level, mobile endpoints, databases, data concentration points, Internet-facing points, cloud objects level, and cloud boundaries level. NIST SP 800-144 standard raised concerns about vulnerabilities in processes and techniques for determining data location, electronic discovery, insider access, composite services, data visibility, defining ownership boundaries, virtual network protection, virtual machine image management, authentication and access control, hypervisor management, data processing/sanitisation, and client-side processes. The standard has made high-level recommendations on IT governance, compliance, trust, architecture, identity management, access control, software isolation, data protection, availability, and incident response. Their recommendations comprise a high emphasis on knowledge-based discovery and decision-making. In this context, the proposer wants to study how data marts can be deployed and used on clouds for implementing knowledge-driven security applications.

Academic researchers have studied about use of data marts in security applications to a reasonable extent. However, deploying such security applications on cloud computing environment is still under-researched. In past few years, researchers have shown significant interest in data mining on the cloud. The proposer wants to extend the existing theories formed on data mining on the cloud to security applications.

Data mining enables knowledge-based discovery of past signatures, traces, patterns, and trends useful in making informed decisions about containment and prevention of attacks[31]. Traditional database based security applications, like anti-malware, intrusion detection systems, and Botnet detectors, make use of ever-increasing records in databases for detecting and suppressing malicious codes and exploits. However, ad-hoc decision-making about containment and preventing attacks may cause false positives and false negatives. This is where data mining can be very useful given their proven role in deploying knowledge-based analytical systems. As prevalent in traditional data mining systems, a rule base is needed for extracting, transforming and loading data from various sources for creating a centralised data warehouse for running analytics. In security systems, the rules need to be defined for detecting abnormal trends and patterns, detecting anomaly signatures in inbound traffic, detecting system misuse, and detecting data breaches [29]. The decision-making on attack containment and prevention of attacks may be based on qualitative analysis by IT security experts supported by lightweight statistics [30]. Some of the indicators of attacks proposed by [30] are SYN counts per second, IP address counts per second, TCP flags per second, average packet size and frequency variations per second, and traces of UDP flooding. Data mines can be used to store such data longitudinally over a period such that a pattern of such attacks can be established and effective containment and prevention strategies can be formulated.

2. SIGNIFICANCE

This research will result in a novel security management framework by employing data mines to detect, contain and prevent attacks on cloud computing systems. Given that the security frameworks and related controls on cloud computing are still evolving, this research may prove to be a useful piece of contribution in enhancing the theoretical foundation established by existing studies. The results from literature studies will be modelled on OPNET simulation tool such that the behaviour, performance and effectiveness of the framework can be observed in a simulated environment. The results will be validated by IT security experts in India such that the recommendations for practical implementation will be presented.

3. PROPOSED AIMS AND OBJECTIVES

The following aims of the dissertation are proposed:

(a) To research the literatures and derive theoretical knowledge about cloud computing security and proposed controls
(b) To study the fundamentals of data mining for security applications and position them in the context of cloud computing with the help of existing studies on data mining deployment on the cloud
(c) To design, model and simulate a security architecture employing applications for attack detection using data mining on the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

cloud, analyse the results, and get the architecture ratified by IT security experts working in India.

The following objectives of the dissertation are proposed:

(a) To study and analyse how security threats are different on cloud computing systems and how they can impact businesses using cloud hosted IT resources
(b) To study and analyse how data mining can be used for detecting attacks on IT systems, containing them, and preventing them
(c) To study and analyse how data mining can be employed for hosting security applications on cloud computing
(d) To model a security architecture for detecting, containing and preventing attacks on cloud hosted systems using distributed data mines
(e) To simulate the model and derive results on behaviour, effectiveness and performance of the model
(f) To interview chosen IT security experts in India for verifying the practical validity of the security architecture and the simulation results
(g) To recommend how this architecture can be deployed on real life cloud computing systems

4. PROPOSED RESEARCH METHODOLOGY AND METHODS

In modern world, academic research studies are conducted under epistemological rules established by positivists and interpreters. Epistemology demands that generation of knowledge and its process should be ratified by communities interested in related fields. Hence, no research study can be conducted in isolation and every study needs to be validated by experts such that the results can be accepted for wider usage in practice or used as a reference for further studies. There are two philosophical approaches for making a research study fit for validation – the Positivism and the Interpretivism approach. Positivism deals with scientific studies following a deductive approach that encompasses application of mathematics and statistics in experimentation, simulation, analytical, modelling, and testing environments for proving theories. It normally adopts quantitative methodology in which the inputs and outputs of the research are in numerical forms only. Interpretivism deals with empirical studies following an inductive approach that encompasses techniques of human interpretations (hermeneutics) for generating new theories. It normally adopts qualitative methodology in which, the inputs and outputs can be in the form of texts and visuals.

In the proposed study, we have chosen positivism as the primary approach for proving a theoretical model of cloud computing security involving data mines in a simulation environment. However, the researcher has also chosen interpretivism approach as a supporting approach for interpreting the validity of the proposed design with the help of inputs provided by experts. Given that the researcher is not a recognised expert in cloud computing security, mixing interpretivism with positivism will help in expanding the opportunities for studying this highly complex science with an open mind and proposing a solution that may be acceptable by interested scientific communities in this field (following epistemology rules). Mixing positivism with interpretivism requires mixing quantitative and qualitative methodologies, which is recognised in academic research as triangulation.

This study will be based on two types of data – secondary data and primary data. Secondary data can be obtained from other similar studies and primary data is the original data that needs to be collected by the researcher conducting a research. In this study, the secondary data will be collected from in-depth literature review of sources like books, journal articles, online and physical library sources, archives, published theses and dissertations, industry reports, press releases, and such other peer reviewed reliable documents.

The primary data will be collected from two different sources – the simulation environment and interviews of experts. The simulation environment will be built on a popular tool called OPNET that is widely used for academic and industrial studies in the networking world[5]. This tool either can generate data with the help of internal event generators or can extract data from a running network using data collection agents. It is unlikely that the researcher will get access to a running production cloud network and hence, using the internal events generator is proposed. The event generator generates data as per the specifications of the components of the model designed with the help of a sound theoretical background. The design needs to be technically accurate for the simulation to run successfully such that the events generator

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

can generate data about the behaviour and performance of each component on the network.

In addition to the primary data collected from the in-built events generator of OPNET, the researcher proposes to collect interview responses from chosen experts in the field of IT security. The interviews will be collected to gain knowledge about validity of the proposed model and its workability in practical scenarios. The researcher is expecting to gain a thorough understanding about practical challenges in using the proposed model and opportunities for conducting further studies in this field.

5. REVIEW OF EXISTING STUDIES

5.1 Data mining for security applications

Data mining has been used for IT security applications requiring inspection and approval/denial decision-making. Some key examples are antimalware, antispam, Internet security servers, intrusion detection, and intrusion prevention. Data mining in these applications help in archiving data about attack like behaviours such that malicious sessions can be separated out of genuine end-customer sessions and appropriate decision-making rules as per security strategies could be formulated, and implemented. The key security objectives in data mining applications for security are:

(a) Establishing validity of an identity claimed by a session requester following a process called authentication.
(b) Authorising the requester with an identity to gain access to resources based on context, predicates, and attributes.
(c) Ensuring that any attempt to modify, tamper, or delete information is detected and blocked (protection of integrity).
(d) Ensuring that information is disclosed to authorised individuals only (protecting confidentiality).
(e) Ensuring that all attempts to disrupt availability of services are detected and blocked (protecting availability)
(f) Ensuring that all activities are logged such that they can be used to generate and enhance knowledge about attack-like behaviours.

[28] described that data mining/warehousing and online analytical processing could help in creating an archive of most relevant information about traffic patterns such that historical analysis of such patterns and knowledge of known attacks could help in detecting attack like behaviours. Attack patterns can be detected by applying association and sequencing rules on the item-sets captured from inbound traffic for detecting attack signatures, length of streams (comparing stream-lengths of valid traffic and known attacks' traffic), and anomaly behaviours. They designed an improved version of Apriori algorithm for supporting length-based decisions taking help of known flow patterns and attack signatures recorded in data mines. [36] described that data mines can help in creating a repository of normal data package behaviours such that any traffic exceeding the limits of the normal data behaviours can be isolated and inspected. Wenjun et. al argued that this method will reduce the workload of intrusion detectors but may increase chances of missing attacks designed to exhibit normal traffic behaviours. Hence, as reiterated by Huang et al. intrusion detection systems cannot be effective by creating static rules. Huang et al. created a data mining model in which the rules are self-learning that modify thresholds based on the knowledge of alerts generated by localised policy servers installed throughout the network at strategic locations. Self-learning rule modelling has been further enhanced by [19] using topology discovery and traffic analysis inputs by a large number of XML based programs acting as network probes. These network probes collect network statistics from nodes and feeding them into a network of integrated data mines, called network information base that in turn is integrated with asset information base for immediate identification of points under attack. Separate log analyser engines are deployed for analysing the records of the network and asset information bases and creating records that help the rules engine to learn on its own. However, dynamic rules engines are resource hungry and need significant expandability in the data mines and number of deployable network probes. Such an expandability is infeasible in many self-hosted environments because of high costs and management overheads.

In practical applications, data mining is used for gathering, organising, and using intelligence information for countering cyber-terrorism, insider threats, malicious intrusions, malware attacks, Internet-based financial frauds, identity thefts, and attacks on critical infrastructures and government facilities. Data mining helps in organised detection and analysis of vulnerabilities and threats, and building their one-to-one, one-to-many and many-to-many ontological mappings such that accurate

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

risk analysis can be done for implementing information security [32]. In addition, system, application, user activity, and administrative/maintenance activity logs could be stored as records in data mines for monitoring the unauthorised activities on sensitive files and predict the possibility of internal or external attacks. Data cubes stored in data mines and exported in the form of XML document-modelling files help in prediction of threats and risks using an integrated activity log analysis framework[1]. Activity logs may comprise of attack signatures collected from various nodes spread across the network such that a collective log analysis could be conducted using data mines comprising records of distributed logs from across the networking infrastructure.

From the above analysis, it is revealed that data mining for security applications poses the following challenges:

(a) Large scale computing and storage capacities for ever growing data volumes
(b) Sophisticated models for data modelling, analysis, and reporting
(c) Distributed data capturing points
(d) Security of data mines

Cloud computing offers major benefits for data mining that may help in addressing these challenges. The next section presents a review of existing literatures on data mining on cloud computing.

5.2 Data mining on the clouds

Data mining on the clouds is a new emerging concept based on traditional technological concepts of data mining in self-hosted infrastructures. As reviewed under "Background and Context", cloud computing comprises massively parallel computing capabilities employing large-scale distributed computing. In addition, cloud computing offers elasticity of resources on demand. These attributes help in improved effectiveness of data mining applications. Hence, data mining on the clouds has better viability and cost effectiveness for business and security applications. [24]. Data mining on the cloud may be correctly described as distributed data mining with multi-agent and multi-tenancy configurations processing large amounts of data employing optimum use of parallelism. Multi-agent software systems comprise distributed and heterogeneous architectures having close mapping with a multidimensional data hierarchy using ontological mappings and faceted search algorithms. In cloud based data mining, the data sources are autonomous (independent of the mining entity), data mining facilitates distributed multi-tasking interactions, comprises of intelligent choosing of sources and data capturing, ensures unlimited scalability by virtue of ever increasing distributed databases, enable collaborative mining, and enables multi-strategy mining from heterogeneous data sources. The rules engine will be primarily based on Apriori and Map-Reduce algorithms for parallel association rule mining of frequently changing data sets. Apriori and Map-Reduce are part of the Hadoop framework. The parallel association rule mining is created in six steps using Map-Reduce and Apriori – partitioning and distributing data and execution tasks, formatting subsets of the data, executing mapping task, operator combining, executing reduce task, and combining the output. Advanced algorithms like neural networks and fuzzy techniques can be employed for multi-dimensional data mapping for sophisticated mining using big data sources.

A schematic of data mining on cloud computing is presented in the figure below [22]:

Journal of Theoretical and Applied Information Technology
20th January 2014. Vol. 59 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

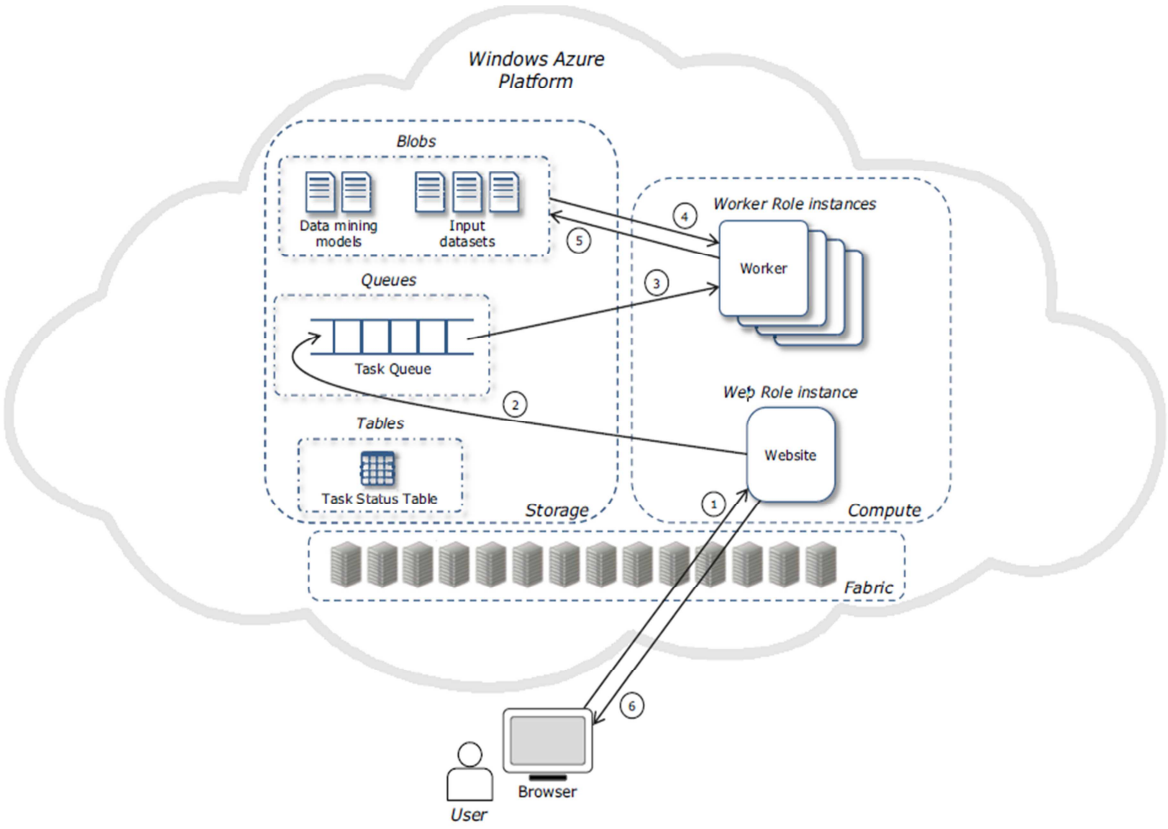

Figure 1: Data Mining On Cloud Computing

The key components of the data mining architecture presented in Figure 1 are blobs (a set of data mining models having their own identifiers and the input data sets), queues, worker role instances, tables, and web role instances. Additional modules are required for mining data storage using data block rearrange algorithms and data block migration algorithms. The key data relationships involved are mining process relations and mining block relations. The process comprises the following steps:

(a) The inputs from the browser are input data location, data mining algorithm identifier, and parameter values.

(b) A set of tasks are entered into the task queue based on the browser inputs.

(c) The task allocator verifies which workers (computing resources) are idle and then allocates tasks from the task queue to each idle worker that is executed at the virtualised servers. The inputs about datasets are based on the browser inputs.

(d) The files needed to capture data and complete the mapping are transferred from the blob to virtual storage on the clouds.

(e) The results are stored back on the blob.

The focus of this research is on security as a service on cloud computing using data mining. Further to the review of data mining approaches in cloud computing, their applications in security as a service are reviewed in the next section.

5.3 Security As A Service On Cloud Computing Using Data Mining

A brief review of cloud computing security is presented under "background and context". This section is dedicated to investigating use of data mining in implementing service oriented security applications on cloud computing. The primary systems to secure are service-oriented architecture, virtualisation, networking components and links, databases, applications, data storage, and computing resources[10].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Detection of anomalies and attacks is carried out using association rules in intrusion detection systems based on collaborated knowledge of attack signatures, anomalies detected and unauthorised activities [24]. The association rules engine on cloud data mining for establishing multidimensional data hierarchy can be employed for security applications on cloud computing. The association rules-based engines need to be automated employing advanced algorithms, like a modified Apriori algorithm employing fuzzy, genetic, neural networks, and decision trees algorithms

The data mining schemes and protocols used in traditional applications (like, SVM, K-Means, and KNN) needs to be modified keeping in mind that cloud servers are shared by a large number of tenants and hence need to be treated as semi-honest. This means that security algorithms (intrusion and anomaly detection) need to be combined with trust algorithms (request, access, identity, ownership, delegation, and notification). The query processing and execution on cloud data mines should be implemented through encrypted schemes handshake between querying client and query processing and execution engine. A separate database for security and authorisation needs to be incorporated for authenticating and authorising users requesting SOA resources through SOAP messages[37. The activities of each authorised user need to be traced and mined at various layers for identifying insider threats. The key layers in the data mines may comprise messaging logs, policy breach logs, conversation logs, delegation logs, audit logs, and exploit attempt logs.

In addition to mining of anomalies, intrusions, and user activities, there should be privacy-preserving data included with the records mined from the main cloud databases. The association algorithms need to be strengthened with classification and correlation principles and techniques for masking, encryption, and distribution in the multiparty computing and storage environment on cloud computing. Each data mine record needs to be associated with data ownership records in the authentication and authorisation databases. This will ensure that the data mines do not store records with anonymous ownership that itself is a prominent security threat on cloud computing.

The challenge is not only of mining relevant records for predictive security analysis of cloud computing. The challenge is also related to processing and analytics of mass log databases employing security analytical applications. Figure 2 presents a schematic of collection of logs, mining of logs, and analysis of logs needed on cloud computing. The association rule-based mining (Apriori algorithm) combined with Map Reduce is currently a valid solution to mining mass-scale logs from SaaS databases. However, applications for analysis of the logs are open for research. Currently, the key challenges in designing log analysis applications are response times, usage functions, parallel processing of queries, dynamic updating of data mines, and securing the data mines

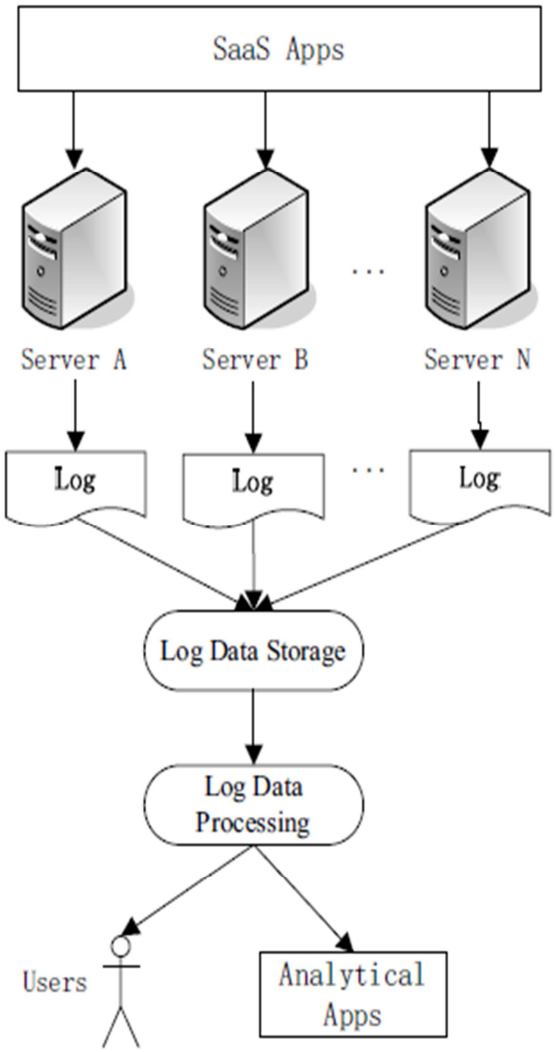

Figure 2: Processing And Analytics Of Mass Log Data On Cloud Computing

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.4 Summarisation

Data mining has been an innovative and productive system in self-hosted infrastructures. However, its growth and popularity is limited by its never-ending demands for significant computing and storage resources. Cloud computing offers a new lease for data mining in variety of applications. Given the high usefulness of data mining in intrusion detection, anomaly detection, and user activity monitoring, an innovative era of cloud security using data mines is emerging. It will be more powerful and sustainable given the unlimited scalability, massive parallel processing capability, and resources elasticity on demand offered by the clouds. The literatures reveal usefulness of integrating traditional data mining algorithms with Hadoop Map-Reduce framework, although research studies are still ongoing. This research will focus on the fundamentals of data mining for security applications and positioning them in the context of cloud computing with the help of existing studies. In addition, this research will present a novel data mining-based security application on cloud computing by designing, modelling, and simulating a security architecture employing applications for attack detection using data mining on the cloud. The results will be analysed and the architecture and its simulation results will be ratified by IT security experts working in India. The framework is expected to value add to the existing body of knowledge on data mining security and security applications using data mining on cloud computing.

REFERENCES:

[1]Alampalayam.S.P. and Kumar, A.(2004)."Predictive Security Model using Data Mining", IEEE Communications Society: p. 2208-2212.

[2] Alborz, N., Keyvani, M., Nikolic, M. And Trajkovic, L. (2000). “Simulation of Packet Data Networks using OPNET”. ACM, p. 1-6.

[3] Bisong, A. and Rahman, S. M. (2011). "An Overview of the security concerns in enterprise cloud computing". *International Journal of Network Security and its Applications (IJNSA)*, Vol.3 (1), p. 30-45.

[4] Carroll, M., Merwe, A. and Kotze, P. (2011). "Secure Cloud Computing: Benefits, Risks and Controls". IEEE, p. 1-9.

[5] Chang-Bin, J., and Li, C. (2010). "Research on Privacy Preserving Data in Web Log Mining", IEEE: p. 1-4.

[6] Chang, X. (1999). “Network Simulations with OPNET”. IEEE: 307-314.

[7] Charlton, S. (2009). "Model-driven design and operations for the cloud". Book chapter: *Towards best practices in cloud computing*, p. 17-26. Eds: Skar, L. A., Lennon, R. and Berre, A. J., proceedings of workshops on cloud computing at OOPSLA09, October 2009 at Orlando, Florida, USA.

[8] Cheung, D. W. (2011). "Security on Cloud Computing, Query Computation and Data Mining on Encrypted Database", IEEE: p. 1-6.

[9] Du, H., and Li, Z. (2011). "Data Rearrange based on Mining Block Access Sequence in Cloud Storage", IEEE Computer Society: p. 2507-2511.

[10] Farroha, B. S. and Farroha, D. L. (2012). "Architecting Security into the Clouds: An Enterprise Security Model", US Department of Defense: p. 1-7, IEEE Xplore.

[11] Guangjuan, L., Ruzhi, X., Ziangrong, Z., and Liwu, D. (2009). "Information Security Monitoring System based on Data Mining", IEEE Computer Society: p. 472-475.

[12] Hadar, E. and Gates, C. (2009). "Cloud Computing Web-Services Offering and IT Management Aspects". Book chapter: *Towards best practices in cloud computing*, p. 27-39. Eds: Skar, L. A., Lennon, R. and Berre, A. J., proceedings of workshops on cloud computing at OOPSLA09, October 2009 at Orlando, Florida, USA.

[13] Huang, N., Kao, C., Hun, H., Jai, G., and Lin, C. (2005). "Apply Data Mining to Defense-in-Depth Network Security System", IEEE, p. 1-4.

[14] Hu, T., Chen, H., Huang, L., and Zhu, X. (2012). "A Survey of Mass Data Mining Based on Cloud computing", IEEE: p. 1-4.

[15] Jansen, W. and Grance, T. (2011). "Guidelines on Security and Privacy in Public Cloud Computing", Special Publication 800-144, National Institute of Standards and Technology (NIST): p. 4-80, U.S. Department of Commerce.

[16] Kaufmann, V., Badger, L., Whiteside, F., Tong, J., Bohn, R., Chu, S., Hogan, M., Liu, F., Mao, J., Messina, J., Mills, K., Sokol, A, and Leaf, D. (2011). “U.S. Government cloud computing technology roadmap – Volume II”, Special Publication 500-293, National Institute of Standards and Technology (NIST), U.S., p. 23-35.

[17] Khorshed, M. T., Ali, A. B. M. S., Wasimi, S. A., (2012), "A survey on gaps, threats remediation challenges and some thoughts for

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

proactive attack detection in cloud computing", *Future Generation Computer Systems (2012)*, doi:10.1016/j.future.2012.01.006. Elsevier.

[18] Li, L., Yang, D., and Shen, F. (2010). "A Novel Rule-based Intrusion Detection System Using Data Mining", IEEE Computer Society: p. 169-172.

[19] Li, L. and Xiao, D. (2010). "Research on The Network Security Management Based on Data Mining", IEEE Computer Society: p. 184-187.

[20] Li, L. and Zhang, M. (2011). "The Strategy of Mining Association Rule Based on Cloud Computing", IEEE Computer Society: p. 475-478.

[21] Lu, Q., Xiong, Y., Gong, X., and Huang, W. (2012). "Secure Collaborative Outsourced Data Mining with Multi-owner in Cloud Computing", IEEE Computer Society: p. 100-108.

[22] Marozzo, F., Talia, D., and Trunfio, P. (2011). "A Cloud Framework for Parameter Sweeping Data Mining Applications", IEEE Computer Society: p. 367-374.

[23] Miller, M. (2009). "Cloud Computing: Web based applications that change the way you work and collaborate online". London: Pearson Publishing. p. 25-38.

[24] Othmane, B., Sidi, R., and Hebri, A. (2012). "Cloud Computing & Multi-Agent Systems: A New Promising Approach for Distributed Data Mining", *In Proceedings of the ITI 2012 34th International Conference on Information Technology Interfaces*, June 25-28, 2012, Cavtat, Croatia, p. 111-116, IEEE.

[25] Othman, Z. A., and Eljadi, E. E. (2011). "Network Anomaly Detection Tools Based on Association Rules", *In 2011 International Conference on Electrical Engineering and Informatics*, 17-19 July 2011, Bandung, Indonesia, p. 1-7, IEEE Xplore.

[26] Pearson, S. and Benameur, A. (2010). "Privacy, Security and Trust Issues Arising from Cloud Computing". IEEE Computer Society: p. 693-702.

[27] Sabahi, F. (2011). "Cloud Computing Security Threats and Responses", IEEE: p. 245-250.

[28] Singhal, A. (2007). "Advances in information security: Data Warehousing and data mining techniques for cyber security", NY: springer.

[29]Song, C. and Ma, K. (2009). "Design of Intrusion Detection System Based on Data Mining Algorithm", IEEE: 370-373.

[30] Stanford, P. J., Parish, D. J. and Stanford, J. M. (2006). "Detecting security threats in the network core using data mining techniques", IEEE: p. 1-4.

[31] Thuraisingham, B., Khan, L., Masud, M. and Hamlen, K. W. (2008). "Data Mining for Security Applications", IEEE Computer Society: p. 585-589.

[32] Wang, B., and Yang, J. (2011). "The State of the Art and Tendency of Privacy Preserving Data Mining", IEEE: p. 1-3.

[33] Wang, H., Meng, L., Cao, M., and Li, Y. (2008). "Data Mining Application Based On Cloud Model In Spatial Decision Support System", IEEE Computer Society: p. 547-551.

[34] Wang, J. A. and Guo, M. (2009). "Security Data Mining in an Ontology for Vulnerability Management", IEEE Computer Society: p. 597-603.

[35] Wang, J., Wan, J., Liu, Z., and Wang, P. (2010). "Data Mining of Mass Storage based on Cloud Computing", IEEE Computer Society: p. 426-431.

[36] Wenjun, L. (2010). "A Security Model: Data Mining and Intrusion Detection", IEEE Computer Society, p. 448-450.

[37] Yamany, H. F. E., and Capretz, A. M. (2008). "Use of Data Mining to Enhance Security for SOA", IEEE Computer Society: p. 551-558.