International Journal of Civil Engineering and Technology (IJCIET)
Volume 8, Issue 10, October 2017, pp. 1709–1717, Article ID: IJCIET_08_10_172
Available online at http://http://www.iaeme.com/ijciet/issues.asp?JType=IJCIET&VType=8&IType=10
ISSN Print: 0976-6308 and ISSN Online: 0976-6316

ONE-TIER CACHE SYSTEM APPLIED TO DATA MINING TECHNIQUES TO ENHANCE THE INFORMATION SECURITY IN THE CLOUD

N.K. Senthil Kumar
Assistant Professor, Computer Science Department,
Vel Tech University, Chennai, India

M. Uvaneshwari
Assistant Professor, Information Technology Department,
Vel Tech University, Chennai, India,

M. Viswanathan
Associate Professor, Computer Science Department,
Vel Tech University, Chennai, India

K. Amsavalli
Assistant Professor, Computer Science Department,
Anand Institute of Higher Technology, Chennai, India

ABSTRACT

Cloud computing is a model makes it possible to achieve on-demand network access and a shared pool of configurable resources thus the popularity is increasing day by day. The purpose of this research work is to grow up the security of the cloud using the techniques such as data mining with reference to the cache system. The cloud environment must contain mechanics for monitoring, reporting and controlling the above mentioned characteristics. For future research purposes, an Apriori algorithm is applied to the one tier cache system. This principle can be applied by all cloud providers and distributors. The Apriori principle can be further used in one tier cache system which is extended up to several components.

Key words: component; formatting; style; styling; insert (key words).

Cite this Article: N.K. Senthil Kumar, M. Uvaneshwari, M. Viswanathan and K. Amsavalli, One-Tier Cache System Applied to Data Mining Techniques to Enhance the Information Security in the Cloud. *International Journal of Civil Engineering and Technology*, 8(10), 2017, pp. 1709–1717.
http://www.iaeme.com/IJCIET/issues.asp?JType=IJCIET&VType=8&IType=10

1. INTRODUCTION

The Cloud Computing represents a model for network access which does not require the same level of maintenance as a standard organizational network would (Mall & Grance, 2011). The cloud is a combination of networks, management solutions, resources, business applications, and data storage. Clouds are so integrated in our everyday life that most people don't even think about them being used. Examples of everyday Cloud Computing are platforms such as iCloud, Office 365 and Google Drive. Cloud computing is a technology that provides various services at minimal cost. Since cloud computing is such a quickly expanding fielding within IT it is not negligent to say that this network model has drastically changed how we perceive networking today. In addition, it has also drastically changed how businesses, organizations and governments act and function. The evolution of Cloud Computing has brought with it new security challenges which is why individuals and businesses alike act hesitant when confronted with the possibility of implementing a cloud solution. Cloud providers, such as Google and Microsoft, might adopt various data analysis techniques to extract valuable information from huge volumes of user data. They use these techniques to identify users' behaviors based on their search history analysis [1, 2]. the different deployment methods the cloud service models can have implications for a clouds security state, it is therefore important to have knowledge of these cloud service models:

Platform as a service (PaaS), infrastructure as a service (IaaS), and software as a service (SaaS). The multinational firms like Google, Amazon and Microsoft provides cloud computing facilities to different sectors. The clouds have computing resources and are delivered through internet. Organizations avoid constructing their own information technology infrastructures. Rather, they are provided with a substitute for hosting their data on the third-party system [3].

According to Wang et al. [4] and Van Wel and Royakkers [5], data mining is the process which examines the data from various perspectives and converts it into useful information. It is majorly used in business applications and Commerce. In addition, data mining is an essential component for gaining huge knowledge. It is usually applied to extract information and patterns understood by humans. Cloud computing providers use data mining to avail inexpensive services. Individuality and privacy are violated by sharing the data and the user is unaware of it and becomes an ethical issue. Thus there is a threat to the cloud user by a hacker. The hacker should not be authorized to the cloud. Simultaneously, he does not have the chance to mine the data of the cloud. In both cases, hackers look for raw and cheap computing power given by cloud computing to mine data and obtain required information from the data.

2. PROBLEM AREA

Cloud computing continues to grow exponentially it is of utter importance to stay updated on the present risks 2 and solutions in order to ensure that future clouds are developed based on the latest insights. The problem presents itself in the massive amount of relevant literature you would have to sift through in order to gain these aforementioned insights. Cloud computing has popularity because of its characteristics multitenancy, scalability, minimized maintenance, and hardware cost Cloud computing poses, complicated privacy and security issues and with relevance to individuals and organizations, large and small, with a need for an up to date security overview as well as a tool for future research. In addition to this, the literature supports the claim as it is consistently mentioned that businesses, which would benefit greatly from a transition to a cloud-based solution, hesitate when faced with the lack of concrete

answers regarding the security of the solution is achieved through techniques of data mining, with specific reference to a one tier cache system.

2.1. Objective

The objectives are as follows:

(i) Cloud computing is a palpable problem thus security threats are identified.

(ii) With one tier Cache system the security of cloud is enhanced using data minig technique.

(iii) To provide essential computing solution to improvize the security of the cloud through data mining techniques.

3. BACKGROUND

Ensuring the confidentiality, integrity and availability, also known as the CIA triangle, has been considered the industry standard for quite some time now (Whitman and Mattord, 2009, p. 8). While these characteristics remain of utmost importance, the rapid development surrounding the IT sector means this definition must be expanded in order to encapsulate the new security situation. Information and communication technology (ICT) can arguably be considered a subcomponent of information security since information security includes the protection of underlying resources. ISO/IEC 13335-1 (2004) defines ICT security as all aspects relating to defining, achieving and maintaining the confidentiality, integrity, availability, no repudiation, accountability, authenticity and reliability of information resources. Thus, von Solms & van Niekerk (2011) argues that a clear understanding of these additional characteristics is essential as without them, information cannot be considered secure. As such, whenever this review utilises the terms “security” or “information security”, this is the definition being referred to.

Virtualization and multi tenancy are two of the core technologies that enable CC to be used as we know it today. A traditional way of hosting applications and data storage involves running one operating system (OS) on one physical server. This traditional hosting method can also be used to create a functioning but inefficient cloud. This is achieved by linking multiple servers using a Virtual LAN (VLAN). This is secure but inefficient in the long term as a large part of the physical hardware available end up being unused. Virtualization was created in order to solve this efficiency problem. By using a Virtual Machine Monitor (VMM) a single physical server can host multiple instances of an OS. This means that a single server can utilise the available hardware power in a more efficient manner (Srinivasan, Sarukesi, Rodrigues, Manoj, & Revathy, 2012). It is also important to note that an emulated OS is still at risk from attacks that targets the traditional version of the OS. For instance a virtual machine running Windows is still at risk from attacks that target normal Windows machines. It is also important to note that hypervisors are additive to the overall security risk (Mishra et al., 2013). Another issue that might have a very severe negative impact on the organisation using a cloud computing solution is data leakage. Data leakage occurs due to the shared resources used by the VMs. These can have the form of cache based attacks or RAM based attacks (Tari, 2014). These attacks occur since both the shared cache and RAM does not automatically flush upon completion of a computing task. To combat the RAM based attacks it is necessary to restrict a VMs ability to lock the memory bus. Both of these solutions requires no expensive hardware modifications but can simply be introduced by adding software. The risks associated with multi tenancy described above have slightly different implications depending on which service model is being used. Data mining algorithms offer solutions for identifying and isolating data security attacks. Such attacks may range from information leakage to fraud and infringement.

Bhagyashree Ambulkar and Vaishali Borkar gives information on data mining systems that have been developed to data for clusters, distributed clusters and grids have assumed that the processors are the scarce resource, and hence shared. It provides result in making better and more efficient decisions. It is judge that mining the data in cloud computing permits service providers to centralize software management and the store data. This outcome gives us the assurance of secure, reliable, and efficient client services.

Astha Pareek and Manish Gupta discussed that Cloud Computing is a web-based technology in which the resources are provided as shared services. The large volume of business data can be stored in Cloud Data centers with low cost. Both Data Mining techniques and Cloud Computing helps the organizations to gain maximized profit and achive low costs in different possible ways. This research made use of the technique of secure replication for building a reliable and secure distributed storage. Thippa Reddy et al. [12,23] developed a idealized framework for security management by deploying data for encountering and preventing security threats on cloud computing systems. The results of the study were estimated and the architecture and its simulation outcomes were verified by information security experts. Sharma and Mehta [13,23] developed an improvised distributed architecture to reduce the challenges and enhance the security of the cloud using data mining. The paper discuss about the data security countered in day to day scenario. However, a distributed architecture was developed for removing such threats found in cloud computing. At the same time, the authors noted that overheads were observed in the system. Therefore, the concept of cache organization was deployed by creating frequent data sets with the help of tools related to data mining. Thus, it can be observed that the cache memory concept was implemented to eliminate the security issue crisis identified in the cloud computing scenarios. Zissis and Lekkas [14] developed a new improvised cloud computing principles for controlling security issues. They adopted software engineering and information system design approaches. The study illustrated that security in a cloud environment requires a systemic point of view, from which security will be constructed based on trust, mitigating protection to a trusted third party. Kumar et al. [15] discussed encryption techniques. The author deployed data compression to utilize unknown unique keys, at the main server level when uploading fully fledged information to the clouds. Khorshed et al. [16] focused on types of security threats that are seen in the cloud using a support vector machine. Bhadauria et al. [17] discussed security issues at various levels through virtualization, application, and network in the cloud computing environment. Their study shared details about security frameworks on the basis of the mechanism of a one-time pass key is deployed. Apart from this, the authors developed data on different security protocols that lies in their concept, and this gives adhered security to users and service providers in a highly contradicting cloud computing. Research by Sasireka and Raja [18,23] deployed an approach to enhance the privacy of cloud environment data by safeguarding it from data mining-oriented threats. The developed approach adopts multiple cloud data distributors and providers as observed. These cloud computing distributors and providers perform fragmentation, data categorization and distribution. Thus in such a framework, cloud providers are not known about the identity of the user. However, date restoration from the clouds was a very complicated task and time-consuming.

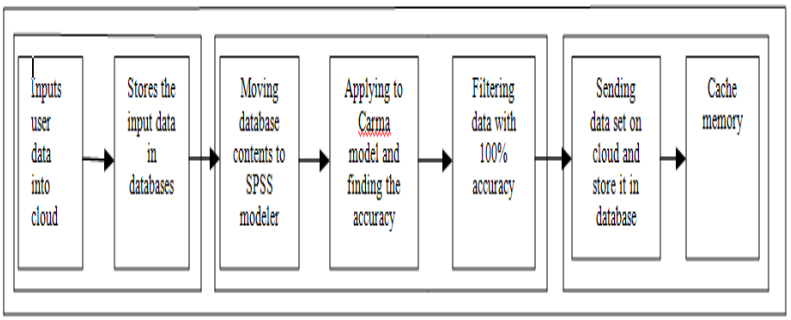

Figure 1 Distributed architecture for the cloud provider

4. IMPLEMENTATION

The two aspects of cloud components are the providers and distributors of the organizations. Distributors of cloud provide data of users in the form of file. These files are broken into blocks, and then they are distributed to different providers in the cloud. Apart from this, the blocks of data are storeh acts like a cache. The cloud provider uses the one tier cache to answer to the queries of the distributor and provides data than searching through the full block of data, which is a real time consuming. Therefore, the file which is accessed regularly performs as cache memory, thus increasing the distributed architecture accuracy, efficiency, as deployed in Figure 1. The end user does not communicate directly with the cloud providers, rather than talk with cloud distributor.

This research proposal proposes the one tier cache memory concept in the distributed architecture for a cloud environment. Distributed architecture offers end users better security and handle all of their cloud data. Whenever a end user gives information to a cloud, large data blocks are created and stored by various cloud providers of different users. If the user needs to perform various global challenges involving all data blocks frequently, this will result in performance overflow. This degrades the performance as system access data. In order to regularize the occurred overhead, the proposed system model checks the data blocks. The information is transmitted to a data mining tool to develop association rules that guides in determining frequent sets of different items that have 100 percent accuracy. This is achieved using a Carma model, an Apriori algorithm, and different methods. These are checked with the association rules in the data mining in relational databases and large transactions of the data [19, 20]. The target fields are not needed in the Carma model. This help us to make the algorithm work in a same way to the construction of an Apriori model. Authors have the freedom to select the items which are listed for Scientific Programming for consequents by redefining the model after it has been created, for example, by the use of user browser which helps to locate either the list of products or the services (antecedents) whose nature is the item provider that you want to popularize. SPSS modeler is a data mining workbench used to synthesis of organized numerical data to create results and make future predictions that gives us the predictive intelligence to make decision on creating the business. The new effective strategies are created by using predictive intelligence and adds data to the cloud distributor to analyze the trend and provide full fledged future interpretation. For example, public sector organizations uses SPSS modeller tools to predict their capacity of the workforce and take immediate measures to maintain public safety and security issues. Additionally, the tools can easily extract and find the personal opinions from text in more than 20 languages and build to better results as their outcomes. Data sets that are used here in most of the baskets are assumed to be frequent occurred. To have a formal definition, assume that there is a number T which supports the threshold. If S is an data set, for which S is the basket number of subset.

Assume S is most frequently occurred, which means its support T or greater value for T . Most occurred data sets are denoted as an "if-then" set of rules. Association rules form $S \rightarrow R$, where S is the data set and R is an item. An association rule's implies that if all the S items are in different basket, then R is likely to be seen in the frequently used basket. Whatever, a notion is likely could be formalized by explaining rule emphasizes $S \rightarrow R$ to be the support ratio for $S \cup \{R\}$. Rule of the basket's fraction with all of S that mixes with R. To develop the cache memory, the following steps were implemented: (1) The end User information is stored in a cloud database. (2) A data file was imported by a data mining tool (SPSS modeler) in an Excel format. (3) the Carma modeling was selected to deploy the most occurring set of items and association rules. (4) Data were filtered for acquiring the threshold values with those have 100 percent accuracy. (5) Information with 100 percent accuracy and efficient was then posted in the cloud environment.

```
private void talashTestAddRemoveObjects() {

    // Test with timeToLiveInSeconds = 200 seconds
    // timerIntervalInSeconds = 500 seconds
    // maxItems = 6
    TalashInMemoryCache<String, String> cache = new TalashInMemoryCache<String, String>(200, 500, 6);

    cache.put("Reliance", "Reliance");
    cache.put("Skrill", "Skrill");
    cache.put("Yahoo", "Yahoo");
    cache.put("Oracle", "Oracle");
    cache.put("HP", "HP");
    cache.put("Airtel", "Airtel");

    System.out.println("Six Cache Object was Added... cache.size(): " + cache.size());
    cache.remove("HP");
    System.out.println("1 object was removed..... cache.size(): " + cache.size());

    cache.put("Tata", "Tata");
    cache.put("Birla", "Birla");
    System.out.println("2 objects were Added but it was reached to maximum Items.. cache.size(): " + cache.size());
```

Figure 2 Inserting data objects in cloud

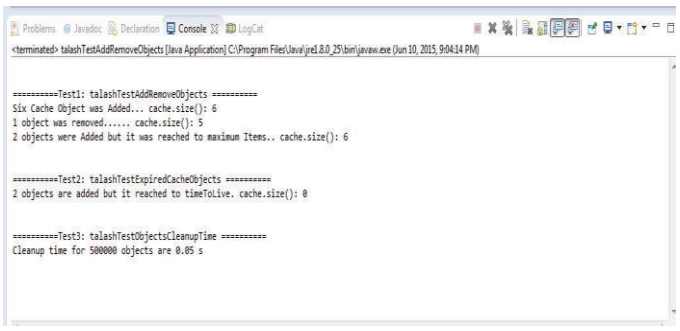

Figure 3 Deletion of data objects in cloud

5. DISCUSSION

It is highly impossible to develop a hundred percent secure network that is secured from threats and alarms. As here and there threats and compromises are being created every day, system developers must drastically improvise their alternative measures to keep data privacy and security. The greater advantage offered by cloud computing, more and more organizations chose to utilize their services to improve performance and low cost. However, cloud providers are prone to the attacks and security threat issues from providers and distributors [1, 3, 4].

The security buglers could use multiple computing techniques to extract information about the user from the hidden data in the cloud. This research proposes distributed cloud architecture to increase security and remove the effect of such

Malicious bugler attacks. However, this will result in overflow since users who require accessing certain data components very often. Thus cache memory concept was implemented in the proposed system by generating often data sets using data mining tools. Even intruder checks the provider data, only one block of data will be exposed to them. Even though this architecture improves data security, it has a considerable amount of overhead issues, if the user decides to access the whole data set very often problem might arise. To overcome this limit, data mining tools are used to identify the most often used data blocks, which is then be stored in a temporary cache memory of the provider [2, 9].

In the proposed research distributed architecture deploys improvised data privacy and security in the cloud. This is marked through having multiple accesses on the targets, or cloud providers, instead of just on only one. This will surely add attackers to utilise more time and effort on gaining access. Additionally, the system has limited data to be checked. Thus intruders will have lees or incomplete data on accessing. Adding the idea of a one tier cache will ensure greater data availability and accessibility. Figure 2 is a snapshot of the code showing the data objects added on the cloud. Six cloud provider items were utilized including HP, Oracle, Skrill, Yahoo, Reliance, and Aircel to store multiple data blocks. The six items are stored in the temporary cache memory for frequent access requests the providers and after an amount of time the stored objects data would be removed from the cache. Figure 3 is a snapshot of the code showing the deletion of objects data. When the six objects were inserted in the cache, and not stored in database. Storing data on cache memory was a temporary task to perform necessary computational analysis operation. A time interval limit was created for the purpose of testing weather deletion of data blocks are done or not. Thus data block would be automatically deleted.

6. CONCLUSION AND FUTURE WORK

The novel approach of this research is to enhance the privacy and security of the cloud through data mining techniques with major rule of using one tier cache system. This proposal is limited to particular cloud computing applications and the research specifies on the use of a one tier cache system. The proposed research techniques are focused on enhancing security of the cloud through data mining techniques. As given detail in the research, e-commerce sites require a cloud server and store multiple often data sets related to the end user who checks in their websites. The Web applications are combined with the cloud services given. In addition, the cloud databases are updated frequently by the imported user. Thus thereafter, the cloud exports add to the database in the SPSS modeler, which then uses Carma modeling, then enhances identifies and accuracy filters data or information with 100 percent confidence. It observed, with the one tier cache system, the security of the cloud application would be improvised to a greater extent. In future work, an Apriori algorithm will be used on one tier cache system cloud providers, vendors, and data distributors. Further, the one tier cache system can be implemented to include 18 objects, on the database-to-SPSS modeller.

REFERENCES

[1] Chen, D., & Zhao, H. (2012). *Data Security and Privacy Protection Issues in Cloud Computing*. In 2012 International Conference on Computer Science and Electronics Engineering (ICCSEE) (Vol. 1, pp. 647–651). http://doi.org/10.1109/ICCSEE.2012.193

[2] Sabahi, F. (2011). *Cloud computing security threats and responses*. In 2011 IEEE 3rd International Conference on Communication Software and Networks (ICCSN) (pp. 245–249). http://doi.org/10.1109/ICCSN.2011.6014715

[3] Trivedi, K., & Pasley, K. (2012). *Cloud Computing Security* (1st ed.). WebEx Communications.

[4] L. Hao and D. Han, "The study and design on secure-cloud storage system," in *Proceedings of the International Conference on Electrical and Control Engineering (ICECE '11)*, pp. 5126–5129, Yichang, China, September 2011.

[5] S. Gupta, S. R. Satapathy, P. Mehta, and A. Tripathy, "A secure and searchable data storage in cloud computing," in *Proceedings of the 3rd IEEE International Advance Computing Conference (IACC '13)*, pp. 106–109, IEEE, Ghaziabad, India, February 2013.

[6] H. Dev, T. Sen, M. Basak, and M. Eunus Ali, "An approach to protect the privacy of cloud data from data mining based attacks," in *Proceedings of the 2012 SC Companion: High Performance Computing, Networking Storage and Analysis (SCC '12)*, pp. 1106–1115, 2012.

[7] J. Wang, J. Wan, Z. Liu, and P. Wang, "Data mining of mass storage based on cloud computing," in *Proceedings of the 9th International Conference on Grid and Cloud Computing (GCC '10)*, pp. 426–431, Shanghai, China, November 2010.

[8] L. Van Wel and L. Royakkers, "Ethical issues in web data mining," *Ethics and Information Technology*, vol. 6, no. 2, pp. 129–140, 2004.

[9] Q. Yang and X. Wu, "10 Challenging problems in data mining research," *International Journal of Information Technology and Decision Making*, vol. 5, no. 4, pp. 597–604, 2006.

[10] L. Torgo, *Data Mining with R: Learning with Case Studies*, Chapman & Hall/CRC, NewYork, NY,USA, 2010.

[11] S. Sharma, A. Chugh, and A. Kumar, "Enhancing data security in cloud storage," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, no. 5, pp. 2132–2134, 2013.

[12] P. Aggarwal and M. M. Chaturvedi, "Application of data mining techniques for information security in a cloud: a survey," *International Journal of Computer Applications*, vol. 80, no. 13, pp. 11–17, 2013.

[13] A. S. Patil, "A review on data mining based cloud computing," *International Journal of Research in Science and Engineering*, vol. 1, no. 1, pp. 1–14, 2014.

[14] S. Singh and R. Sapra, "Secure replication management in cloud storage," *International Journal of Emerging Trends and Technology in Computer Science*, vol. 3, no. 2, pp. 251 254, 2014.

[15] G. Thippa Reddy, K. Sudheer, K. Rajesh, and K. Lakshmanna, "Employing data mining on highly secured private clouds for implementing a security-asa- service framework," *Journal ofTheoretical and Applied Information Technology*, vol. 59, no. 2,pp. 317–326, 2014.

[16] S. Sharma and H. Mehta, "Improving Cloud Security Using Data Mining," *IOSR Journal of Computer Engineering*, vol. 16, no. 1, pp. 66–69, 2014.

[17] D. Zissis and D. Lekkas, "Addressing cloud computing security issues," *Future Generation Computer Systems*, vol. 28, no. 3, pp. 583–592, 2012.

[18] A. Kumar, H. Lee, and R. P. Singh, "Efficient and secure cloud storage for handling big data," in *Proceedings of the 6th*

[19] *International Conference on New Trends in Information Science and Service Science and Data Mining (ISSDM '12)*, pp. 162–166, Taipei, Taiwan, October 2012.

[20] M. T. Khorshed, A. B. M. S. Ali, and S. A.Wasimi, "A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing," *Future Generation Computer Systems*, vol. 28, no. 6, pp. 833 851, 2012.

[21] R. Bhadauria, R. Borgohain, A. Biswas, and S. Sanyal, "Secure authentication of cloud data mining API," *Acta Technica Corviniensis-Bulletin of Engineering*, vol. 3, no. 1, 2014.

[22] K. Sasireka and K. Raja, "An approach to improve cloud data privacy by preventing from data mining based attacks," *International Journal of Scientific and Research Publications*, vol. 4, no. 2, pp. 1–4, 2014.

[23] [22] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco, Calif,USA, 2006.

[24] J. Han, "Data mining techniques," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '96)*, p. 545, Montreal, Canada, June 1996.

[25] M. Uvaneshwari and N.K. Senthil Kumar, 2017. Load Balancing and Runtime Prediction using Map Reduce Framework. *International Journal of Civil Engineering & Technology (IJCIET)* - Scopus Indexed.Volume:8,Issue:10,Pages:834-842.

[26] Dr. E V Ramana, S Sapthagiri and P Srinivas, Data M ining Approach for Quality Prediction and Improvement of Injection Molding Process through SANN, GCHAID and Association Rules. *International Journal of Mechanical Engineering and Technology* , 7(6), 2016, pp. 31–40.

[27] Dr. E V Ramana, S Sapthagiri and P Srinivas, Data Mining Approach for Quality Prediction of Injection Molding Process Through Statistica SVM, KNN and GC & RT Techniques. *International Journal of Mechanical Engineering and Technology*, 7(6), 2016, pp. 22–30.

[28] Lakshmi. R and Antony Selvadoss Thanamani, Data Mining Based Dynamic Replication Algorithm for Improving Data Availability in Data Grids. *International Journal of Computer Engineering and Technology*, 7(5), 2016, pp. 09–16.