Hindawi Publishing Corporation
Scientific Programming
Volume 2016, Article ID 2385654, 5 pages
http://dx.doi.org/10.1155/2016/2385654

Research Article

Applying Data Mining Techniques to Improve Information Security in the Cloud: A Single Cache System Approach

Amany AlShawi

King Abdulaziz City for Science and Technology, P.O. Box 6086, Riyadh 11442, Saudi Arabia

Correspondence should be addressed to Amany AlShawi; aalshawi@kacst.edu.sa

Received 31 January 2016; Revised 5 June 2016; Accepted 15 June 2016

Academic Editor: José L. Vázquez-Poletti

Presently, the popularity of cloud computing is gradually increasing day by day. The purpose of this research was to enhance the security of the cloud using techniques such as data mining with specific reference to the single cache system. From the findings of the research, it was observed that the security in the cloud could be enhanced with the single cache system. For future purposes, an Apriori algorithm can be applied to the single cache system. This can be applied by all cloud providers, vendors, data distributors, and others. Further, data objects entered into the single cache system can be extended into 12 components. Database and SPSS modelers can be used to implement the same.

1. Introduction

The cloud is a combination of networks, management solutions, computing resources, business applications, and data storage. It supports a new era of information technology and customer service. Cloud computing is a technology that provides various services at minimal cost. It facilitates data storage and provides multiple levels of information security. By adopting cloud services, a user forgoes the additional cost of buying unnecessary computational resources. 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].

Clouds offer three types of services: platform as a service (PaaS), infrastructure as a service (IaaS), and software as a service (SaaS). Most of the major technology firms, like Microsoft, Google, and Amazon, are now providing cloud facilities to different organizations. The cloud involves computing resources, namely, software and hardware. These are delivered as a service through the 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 of examining data from various perspectives and converting it into useful information. It is widely used in economics and business applications. In addition, data mining is an essential component for knowledge discovery. It is usually applied to extract information and patterns understood by humans. Cloud computing providers use data mining to offer users cost-effective services.

If clients are unaware of the data being gathered, then ethical issues such as individuality and privacy are violated. If cloud providers misuse the data, such things could become a serious threat to the privacy of data. An attacker who rests outside the cloud network does not have authorized access to the cloud. Simultaneously, he does not have the chance to mine the data of the cloud. In both cases, attackers could adapt raw and cheap computing power given by cloud computing to mine data and obtain required information from the data [6, 7].

2. Problem

Cloud computing is gaining popularity because of its features, including multitenancy, scalability, minimized maintenance, and hardware cost. Cloud technology gives the client multifold facilities, but it also brings additional privacy and

security issues. Cloud computing poses new, complicated security issues for numerous reasons. Conventional cryptographic primitives used for ensuring data security cannot be used directly. Assured correctness of storage under an update of dynamic data could be provided by the modification, insertion, and deletion of stored data [8]. Therefore, this study intended to focus on enhancing the security of the cloud. This was achieved through techniques of data mining, with specific reference to a single cache system.

3. Objectives

The objectives are as follows:

(i) Identify various security threats in cloud computing.

(ii) Enhance the security of the cloud through data mining techniques by making use of a single cache system.

(iii) Provide valuable suggestions to enhance the security of the cloud through data mining techniques.

4. Background

Dev et al. [3] developed an approach to safeguard cloud data privacy from the mining of database attacks. In their approach, it was stated that assuring the security of data in the cloud is a challenging issue. Cloud service providers and other third parties make use of these unique data mining techniques to obtain valuable information from users about the data stored on the cloud. The research focused on the effect of mining data on the cloud. Authors developed a distributed structure for eliminating mining-based privacy issues related to data in the cloud. The study approach integrated fragmentation, categorization, and distribution. Thereby, it safeguarded mining of the data by maintaining levels of privacy, separating chunks of data, and storing such data chunks for needed cloud providers. Apart from these functions, it was noted that the developed system provides an efficient way to protect privacy from mining-oriented attacks. However, this resulted in performance overhead when users require access to all data frequently. For example, some users might require carrying out a global analysis of data. Such analysis would require data to be obtained from various locations, which will result in degraded performance. It can be concluded that the developed system provided an efficient way to protect privacy from mining-oriented attacks. But it incorporated overhead performance when users require access to the entire data.

Sharma et al. [8] carried out research to enhance data security in cloud storage. This research focused on architectural components to provide data security for administrators and users. For ensuring the correctness of a client's information in the storage of cloud data, this study developed a flexible and effective distributed scheme. The scheme was an explicit support from dynamic data, encompassing block delete, update, and append operations. The study used the data encryption standard (DES) algorithm for information security. This algorithm allows information stored in the database to be seen in the required format. The database may be in the form of cipher text and, at the same time, can be accessed when the data are requested. This phenomenon is dependent on the erasure-correcting code distributing the file. The file is prepared to give vectors in redundancy parity and guarantee the dependability of data. By adopting tokens of homomorphism with erasure code and data distributed verification, the scheme achieved correct storage insurance integration and localization of data errors. That is, whenever corrupt data is tracked during the verification of storage correctness across the distributed servers, the misbehaving server's identification is guaranteed. Thus, it can be understood through the analysis that a distributed scheme was developed for assuring the accuracy of client information.

Aggarwal and Chaturvedi [9] reviewed how data mining techniques and relevant algorithms could play a significant role in ensuring data security in the cloud. The authors stated that, with the increase in humans' dependence on machines, developing a better and more effective framework for providing a secure electronic infrastructure is required. Clouds provide on-demand services at a much more minimal rate with fewer overheads. This practice maximizes the popularity of the cloud among different entities. However, the authors noted that problems with data security became crucial. This includes users' authentication services and data encryption and protection. In addition, they stated that it is essential to deploy data security in such a way that authorized users can obtain the maximum number of services. At the same time, it is important to enable the detection of unauthorized users in order to stop them from disrupting and misusing the cloud services. Data mining algorithms offer solutions for identifying and isolating data security attacks. Such attacks may range from information leakage to fraud and infringement.

Aishwarya S. Patil and Ankita S. Patil [10] reviewed data mining on the basis of cloud computing, which is a significant characteristic of infrastructure. It aids in making better and more efficient knowledge-driven decisions. It was noted that mining the data in cloud computing permits organizations to centralize software management and the storage of data. This resulted in the assurance of secure, reliable, and efficient client servicers.

Singh and Sapra [11] discussed the management of secure replication for storage in the cloud. They discussed general principles of a new approach to carry out secure replication, especially for stored data. The approach was a dominant technique that provides better outcomes for the availability and security of information. This research made use of the technique of secure replication for building a reliable and secure distributed storage. The enhancement carried out in this technique maximizes the quality by using unique data mart hosting with a cloud provider and this technique stores data based on its sensitivity. This technique is useful for financial organizations and cloud provider firms.

Thippa Reddy et al. [12] developed a novel framework for security management by deploying data for detecting, containing, and preventing attacks on cloud computing systems. The outcomes of the study were estimated and the

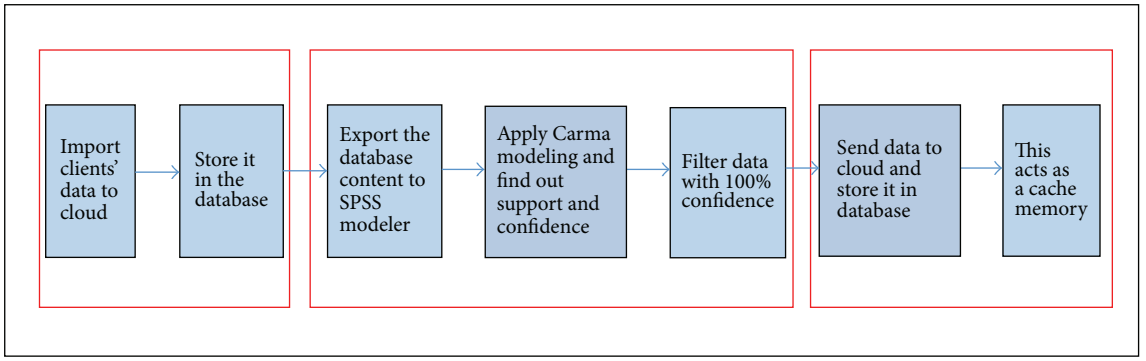

Figure 1: System architecture for a cloud provider.

architecture and its simulation outcomes were ratified by information security experts.

Sharma and Mehta [13] developed an effective distributed architecture to reduce the risks and enhance the security of the cloud using data mining. Authors sated that every day new attacks are discovered and new countermeasures are developed for keeping data secure. Providers and attackers utilize effective data mining techniques to obtain information regarding a client's data stored in the cloud. However, a distributed architecture was developed for eliminating such threats found in cloud computing. At the same time, the authors noted that overheads were seen in the system. Therefore, the concept of cache memory was deployed by creating frequent item sets with the help of tools related to data mining. Thus, it can be concluded that the cache memory concept was implemented to eliminate the threats identified in the cloud computing environment.

Zissis and Lekkas [14] developed unique cloud principles for controlling security threats. 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. They applied data compression to utilize secret keys, especially at the main server level when uploading information to the clouds. Khorshed et al. [16] focused on types of attacks that are found in the cloud using a support vector machine.

Bhadauria et al. [17] discussed security at various levels through application, network, and virtualization in the cloud computing environment. Their study developed security frameworks on the basis of the mechanism of a one-time pass key. Apart from this, the authors found that the uniqueness of developed security protocols lies in their concept, and this gives security to users and service providers in a highly conflicting cloud computing environment.

Research by Sasireka and Raja [18] developed an approach to enhance the privacy of cloud data by safeguarding it from data mining-oriented attacks. It was observed that the developed approach adopts multiple cloud data distributors and providers. These distributors and providers perform data categorization, fragmentation, and distribution. In such a system, cloud providers are not aware of the identity of the user. However, date restoration from the clouds was a very complicated task.

5. Implementation

Cloud providers and distributors of data are two main system components. Distributors of cloud data acquire data in file form from the users. These files are separated into chunks, and then they are distributed across different providers in the cloud. In addition to this, the chunked data are stored by the cloud provider and are often accessed by a client after they are examined. The client stores such information separately in another file, which performs like a cache. The cloud provider responds to the queries of the distributor by providing data from the cache instead of searching through the whole chunk of data, which is a more time-intensive process. Therefore, the file is mostly accessed data that performs as cache memory, thereby maximizing the distributed architecture efficiency, as depicted in Figure 1. Users do not interact with providers in the cloud directly, rather doing so through a cloud data distributor.

This study deployed the single cache memory concept in the distributed architecture. Such architecture offers users better security for all of their cloud data. Whenever a client supplies information to a cloud, data chucks are created and stored by various cloud providers. If the user needs to perform global operations involving all data chunks frequently, this will result in performance overhead. This is due to the fact that the system must access information from different cloud providers resulting in degraded performance. In order to minimize the encountered overhead, the proposed system model examines the data chunks. The information is transmitted to a data mining tool to develop association rules that assist in determining frequent sets of items that have 100 percent confidence. This is achieved using a Carma model, an Apriori algorithm, and other methods. These are referred to as association rules in the mining in relational databases and large transactions [19, 20].

Input or target fields are not required in the Carma model. This is useful to make the algorithm work in a similar way to the construction of an Apriori model. You have the freedom to constrain or to select the items which should be listed

as antecedents or consequents by refining the model after it has been created, for example, by the use of model browser which helps to locate either the list of products or the services (antecedents) whose subsequent is the item that you want to popularize.

SPSS modeler is a data mining workbench used for the analysis of organized numerical data to create outcomes and make future predictions that lend predictive intelligence to business decision-making. Use of predictive intelligence creates effective strategies as it permits the organization to analyze its trends and provide insight to future interpretation. For example, public sector organizations utilize SPSS modeler tools to predict their workforce capacity and take measures to maintain public safety issues. Additionally, the tools can quickly extract and determine personal opinions from text in more than 30 languages and help build more detailed outcomes.

Item sets that appear in most of the baskets are assumed to be frequent. To have a formal definition, assume that there is a number T support threshold. If S is an item set, support for S is the basket number for which S is a subset. Assume S is frequent, which means its support is T or greater. Frequent item data sets are denoted as an "if-then" collection of rules. Association rules form $S \rightarrow R$, where S is the items set and R is an item. An association rule's implication is that if all the S items are in some basket, then R is likely to be seen in the same basket. Moreover, a notion of what is likely could be formalized by explaining rule confidence $S \rightarrow R$ to be the support ratio for $S \cup \{R\}$. Rule confidence is the basket's fraction with all of S that involves R.

To deploy the cache memory, the following steps were implemented:

(1) User information was stored in a cloud database.

(2) A file was imported by a data mining tool (SPSS modeler) in an Excel format.

(3) Carma modeling was selected to develop the frequent set of items and association rules.

(4) Data were filtered for acquiring the values that have 100 percent confidence.

(5) Information with 100 percent confidence was then posted in the cloud environment.

6. Discussion

No one can claim that it is possible to develop a hundred percent secure network that is immune to all types of attacks. As new threats and compromises are being initiated every day, system developers must improve their countermeasures to keep data private and secure. In light of the tremendous benefits offered by cloud computing, more and more organizations chose to utilize their services to improve performance and decrease cost. However, cloud providers are susceptible to attacks and security threats from insiders and outsiders [1, 3, 4].

Attackers could use multiple computing techniques to extract information about the user from the data stored in the cloud. This research proposes a distributed cloud architecture to increase security and minimize the effect of such malicious attacks. However, this will result in overheads since users might require to access certain data components very frequently. Hence cache memory concept was implemented in the proposed system by generating frequent item sets using data mining tools. Even if an attacker gains access to the cloud provider's storage space, only one chunk of data will be exposed. Even though this architecture increases data security, it involves a considerable amount of overhead if the user decides to access the whole data set frequently. To overcome this drawback, data mining tools are used to determine the most frequently used data chunks, which would then be stored in a temporary cache memory [2, 9].

```
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: Adding objects to the cloud.

```
Problems  Javadoc  Declaration  Console  LogCat
<terminated> talashTestAddRemoveObjects [Java Application] C:\Program Files\Java\jre1.8.0_25\bin\javaw.exe (Jun 10, 2015, 9:04:14 PM)

==========Test1: talashTestAddRemoveObjects ==========
Six Cache Object was Added... cache.size(): 6
1 object was removed..... cache.size(): 5
2 objects were Added but it was reached to maximum Items.. cache.size(): 6


==========Test2: talashTestExpiredCacheObjects ==========
2 objects are added but it reached to timeToLive. cache.size(): 0


==========Test3: talashTestObjectsCleanupTime ==========
Cleanup time for 500000 objects are 0.05 s
```

Figure 3: Removing objects from the cloud.

The proposed distributed architecture provides better data privacy and security. This is achieved through having multiple attack targets, or cloud providers, instead of just one. This will surely require attackers to invest more time and effort to be able to gain access. Additionally, the system decreases the amount of data available at each target. So even if one cloud provider's system is compromised, attackers will gain access to incomplete data. Adding the concept of a single cache will ensure greater data availability.

Figure 2 is a snapshot of the code showing the objects added to the cloud. Six cloud provider items were utilized including Yahoo, Reliance, Aircel, Oracle, Skrill, and HP to store multiple data chunks. The six items would be stored in the temporary cache memory to be available for frequent access requests and after a specific amount of time the stored objects would be deleted from the cache.

Figure 3 is a snapshot of the code showing the removal of objects. When the six objects were added in the cache, they were not stored in any database. Storing them in the cache memory was a temporary operation to perform necessary analysis or any other computational operation. A time limit was generated for the purpose of testing to delete those data chunks. The data would be automatically deleted after

a specific time limit. So objects were added in the cache and removed when each reached its maximum item time limit.

For example, if a big online e-commerce site adopting a cloud server needs to frequently store client-visiting data, it could combine the proposed tool and all of its related information with its Web application. That would be achieved through a direct connection with the function "cache.put" as shown in Figure 2. When entering information into the cache, there is a different procedure for each vendor. Vendors have to adopt their own functional application programming interfaces to connect with the cache system.

7. Conclusion and Future Work

The purpose of this research is to enhance the security of the cloud through data mining techniques with specific reference to a single cache system. This study is limited to specific cloud computing applications and the research findings make use of a single cache system. The proposed technique focuses on maintaining security of the cloud through data mining techniques.

As illustrated in the research, e-commerce websites adopt a cloud server and store multiple frequent data sets related to the clients who visit their websites. They combine their Web applications with the cloud services. In addition, they import users' data to the cloud and store them in the cloud database. Thereafter, the cloud exports the contents of the database to the SPSS modeler, applies Carma modeling, identifies support and confidence, and filters data or information with 100 percent confidence. It was observed that, with the single cache system, the security of the cloud application could be enhanced. For future work, an Apriori algorithm can be applied to the single cache system for all cloud providers, vendors, and data distributors. Further, the single cache system can be extended to include 12 objects, which could then be used to implement the database-to-SPSS modeler.

Competing Interests

The author declares no competing interests regarding the publication of this paper.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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 of Theoretical and Applied Information Technology*, vol. 59, no. 2, pp. 317–326, 2014.

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

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

[15] A. Kumar, H. Lee, and R. P. Singh, "Efficient and secure cloud storage for handling big data," in *Proceedings of the 6th International Conference on New Trends in Information Science and Service Science and Data Mining (ISSDM '12)*, pp. 162–166, Taipei, Taiwan, October 2012.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.