A Multi Agent System for Secured Storage and Effective Access in Cloud Using Data Mining Approach

A. Antonidoss and D. Manjula
Department of Computer Science and Engineering, CEG Campus,
Anna University, 600025 Chennai, India

Abstract: The cloud computing storage approaches have been increasingly adopted by several experts from various domains. The cloud computing system interacts with various systems and these services are offered from data centres all over the world. The Multi-Agent System (MAS) includes various data agents that interact with the Cloud Data Storage (CDS) which in turn interacts with the Data Mining Agents. The data mining agents analyses data from different perspectives and summarizes it into useful information. The main process in the distributed environment for providing security and optimized access for data are: the distributed data mining provides the in-depth information of the data and provides interaction with various processing agents in parallel. The publically available security standards are introduced to incorporate cloud security goals, to avoid cloud threats. The distributed data mining systems provides improved access to the data with the analytical view of the data stored in cloud data storage. In this study, new algorithms for finding association rules and sequential pattern mining are proposed for data analysis from cloud databases. For this purpose, the data are encrypted using Hill cipher and they are analyzed and stored in encrypted form. For this purpose, a MAS is used for performing analysis effectively. The experimental results show the performance of the proposed model provides higher security and accurate analysis in comparison with other existing works.

Key words: Cloud Data Storage (CDS), Multi-Agent system (MAS), data mining agents, Cloud Service Provider (CSP), distributed data mining

INTRODUCTION

Now a days, the volume of business/commercial data is drastically increasing, so maintaining this data manually is very challenging task. Handling this data must be well disposed and inventive in a competitive way where decisions based on analysis of these data can be made quickly. Data mining is one of the most important research areas in information systems. The field of data mining is growing with new strategies and approaches to use our information (data) to extract a useful knowledge. Managing these data manually is a tough task, so we are going to use cloud for our storage purpose (Talib *et al.*, 2011).

Cloud computing is the long dreamed vision of computing as a service where users can remotely store their data in the cloud to enjoy on-demand high-quality applications and services. The cloud computing treats resources as unified entity that makes IT enterprise take a step ahead by storing data remotely into the cloud. It provides great convenience to clients like independence from geographical location, reduces the burden in

storage management and also puts down the expenditure of software and hardware equipments. Online storage service providers are named as Cloud Service Provider (CSP). They can also be called as cloud server (Talib *et al.*, 2011).

As internet based storage services are available with increased network and width, they provide a wide range of storage services which reduce the burden of local machine for storing data. Apart from these benefits, the new storage paradigm faces some challenging issues like security, integrity and privacy of the client's outsourced data. The two major types of attacks detected in cloud servers are internal and external attacks. Internal attacks are initiated by malicious and unreliable CSP and External attacks are initiated by motivated attackers. Due to these attacks, correctness of the data is at high risk due to few data loss incidents such as gmail disaster event. To overcome this problem we use erasure correction code. It is an algebraic property performed on distributed data blocks to retrieve the lost data blocks. This encodes, the message of k symbols into a code word of n symbols by erasure coding after encrypting the data by cryptographic

---

Corresponding Author: A. Antonidoss, Department of Computer Science and Engineering, CEG Campus, Anna University, 600025 Chennai, India

method. Multi Agent System (MAS) includes various data agents that interact with the Cloud Data Storage (CDS), decrypts the data and then decodes the data in the cloud data storage. Then it interacts with the data mining agents. The data mining agents uses a mining algorithm to analyse data from different perspectives and summarizes it into useful information (Abdullah *et al.*, 2011; Talib *et al.*, 2011)).

MATERIALS AND METHODS

Literature review: There are many works have been done by various researchers in this direction in the past. Among them, Talib *et al.* (2012) proposed a Multi Agent System (MAS) architecture based on integrity policy for secure Cloud Data Storage (CDS). Their architecture consists of two types of agents, namely Cloud Service Provider Agent (CSPA) and Cloud Data Integrity Backup Agent (CDIBA) for effective data access in cloud environment. CSPA enhance the data access by the system. CDIBA helps to reconstruct the original cloud data using a new technique of backing up the cloud data regularly using Structural Query Language (SQL) programming. An elastic cloud resource allocation mechanism that autonomously and dynamically reallocates cloud resources on demand to BoT executions is proposed by Gutierrez-Garcia and Sim (2013).

Zhang *et al.* (2005) and Ejarque *et al.* (2010) proposed a new approach of the Semantically Enhanced Resource Allocation (SERA) distributed as a multi-agent system. They presented a distributed resource allocation method which is the combination of the benefits of semantic web for making easy integration and agent technologies for coordinating and adapting the execution across the different providers. Zhang *et al.* (2005) discussed about a story of synergism of two cutting edge technologies using agents and data mining. Lin and Tzeng (2012) proposed a multi agent system architecture for providing comprehensive security on cloud data storage.

Abdullah *et al.* (2011) proposed a model of knowledge management system for facilitating Knowledge as a Service (KaaS) in cloud computing environment for effective data access in distributed environment. Lin and Tzeng (2012) proposed a secure erasure code-based cloud storage system for secured data forwarding in distributed cloud environment. Muthurajkumar *et al.* (2015a) proposed temporal secured cloud log management techniques for providing security to maintain transaction history in cloud within time period. Later on they introduced an agent based security mechanism for providing security to the cloud data (Muthurajkumar *et al.*, 2015b).
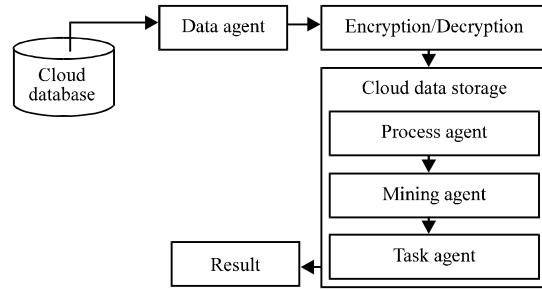

Fig. 1: System architecture

A new multi agent based conjunct model was proposed by Xu *et al.* (2010) in which is dividing the computing environment into three layers namely, adopting agent to represent the behaviour and resource of node to realize more efficient resource sharing and collaboration.

System architecture: The proposed system architecture is shown in Fig. 1. It consists of four major components namely cloud database, data agent, encryption/decryption module, data storage module and the result. Cloud database holds the large volume of data. Data agent extracts the relevant and clean data from the cloud database using data servers. Encryption/Decryption module is secured the retrieved data by using cryptography algorithm in this research.

Cloud data storage module stores the secured data with the help of process agent, mining agent and task agent. Process agent stores the data sequentially based on the proposed sequence analysis algorithm. Mining agent finds and selects the suitable secured data using the proposed frequent pattern matching algorithm. Task agent grouped the data and sent to the result module.

Proposed algorithm: In this study, we propose a new multi agent system for handling the cloud data in distributed environment and parallel processing manner with the help of the two newly proposed algorithms. First algorithm called Sequence Analysis algorithm is for analysing the cloud data in sequence. Second algorithm called frequency pattern matching algorithm for identifying the relationship level of each data in cloud data. The steps of the proposed algorithm are as follows:

Sequence analysis algorithm:
- Step 1: Read dataset and form matrices
- Step 2: Read the key matrix
- Step 3: Apply hill cipher and perform encryption.
- Step 4: for i = 1 to F1 do begin
- Step 5: Generate the candidate sets

- Step 6: For all input sequences s in the database
- Step 7: If support >threshold then increment count
- Step 8: If count >3 then stop
- Step 7: End

This algorithm is used to check the count of sequential patterns up to 3 using Association Rule Mining algorithm.

Frequency pattern matching algorithm:
- Step 1: Read dataset
- Step 2: Find the one item frequent set
- Step 3: Find the support and confidence
- Step 4: Call the decision agent for checking whether it is below threshold
- Step 5: If support>threshold then Find the two item frequent set
- Step 6: Call the decision agent for checking whether it is below threshold
- Step 7: If support>threshold then Find the three item frequent set
- Step 8: Call the decision agent for checking whether it is below threshold
- Step 9: Form association rules
- Step 10: Produce results

This algorithm provides the frequency matching details of the given cloud data. This result helps to access the most relevant information from the large volume of data stored in cloud environment.

RESULTS AND DISCUSSION

The proposed system is implemented using JAVA programming language. The Performance of the system is tested in terms of number of threats received and number of packets transmitted at different time intervals using agents and without using agents. Figure 2 shows the performance analysis between the "Number of threats" received with and without agents at 30 m sec by conducting various experiments with different number of messages sent. From Fig. 2, it can be observed that the performance of the proposed multi agent system is better than without agent system in terms of threat received in particular time period from cloud.

Table 1 shows the comparative analysis between with mining and without mining algorithm used in the distributed cloud system.

From Table 1, it can be observed that the performance of the proposed system (Encryption+FSA+FPMA) is performed well when it is compared with encryption only and encryption with frequent sequence analysis. This is due to the fact that uses of effective data mining algorithm.
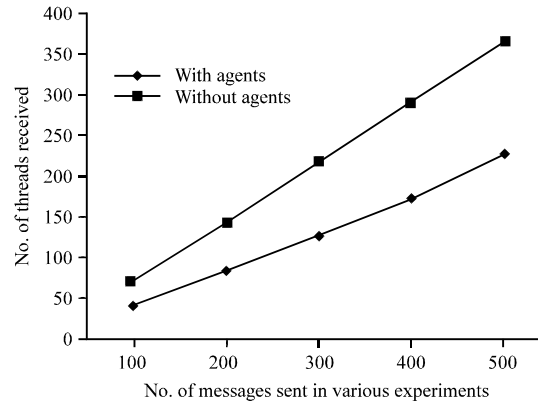

Fig. 2: Performance analysis

Table 1: Comparative analysis

No. of messages sent	No. of threats received		
	Encryption only	Encryption+Frequent sequence analysis	Encryption+ FSA+FPMA
100	55	44	37
200	91	83	70
300	145	128	116
400	179	167	149
500	264	229	214

CONCLUSION

A multi agent system with secure cloud data storage is proposed and implemented in this paper. The proposed Multi Agent System (MAS) includes various data agents and data mining agents. The data mining agents with the mining algorithm such as frequency pattern matching algorithm and Sequence Analysis algorithm, analyse data from different perspectives and summarize it into useful information. The proposed system provides the secured cloud storage for storing our data on cloud. Security goals are implemented to avoid cloud threats. The distributed data mining provides in-depth information about the data and provides interaction with various processing agents in parallel. It also provides improved access to the data with the analytical view of the data stored in cloud data storage. Future works in this direction could be the uses of effective cryptographic algorithm to improve the security of cloud data.

REFERENCES

Abdullah, R., Z.D. Eri and A.M. Talib, 2011. A model of knowledge management system for facilitating knowledge as a service (KaaS) in cloud computing environment. Proceedings of the International Conference on Research and Innovation in Information Systems, November 23-24, 2011, Kuala Lumpur, pp: 1-4.

Ejarque, J., R. Sirvent and R.M. Badia, 2010. A multi-agent approach for semantic resource allocation. Proceedings of the IEEE 2nd International Conference on Cloud Computing Technology and Science, November 30-December 3, 2010, Indianapolis, IN., pp: 335-342.

Gutierrez-Garcia, J.O. and K.M. Sim, 2013. A family of heuristics for agent-based elastic cloud bag-of-tasks concurrent scheduling. Future Generat. Comput. Syst., 29: 1682-1699.

Lin, H.Y. and W.G. Tzeng, 2012. A secure erasure code-based cloud storage system with secure data forwarding. IEEE Transact. Parallel Distributed Syst., 23: 995-1003.

Muthurajkumar, S., S. Ganapathy, M. Vijayalakshmi and A. Kannan, 2015a. An effective intrusion detection on cloud virtual machines using hybrid feature selection and multiclass classifier. Aust. J. Basic Applied Sci., 9: 38-41.

Muthurajkumar, S., S. Ganapathy, M. Vijayalakshmi and A. Kannan, 2015b. Secured temporal log management techniques for cloud. Procedia Comput. Sci., 46: 589-595.

Talib, A.M., R. Atan, R. Abdullah and M. Azrifah, 2011. CloudZone: Towards an integrity layer of cloud data storage based on multi agent system architecture. Proceedings of the IEEE Conference on Open Systems, September 25-28, 2011, Langkawi, pp: 127-132.

Talib, A.M., R. Atan, R. Abdullah and M.A.A. Murad, 2012. Towards a comprehensive security framework of cloud data storage based on multi agent system architecture. J. Inform. Secur., 3: 295-306.

Xu, X.L., C.L. Cheng and J.Y. Xiong, 2010. Conjunct model of cloud and client computing based on multi-agent. J. China Inst. Commun., 31: 203-211.

Zhang, C., Z. Zhang and L. Cao, 2005. Agents and Data Mining: Mutual Enhancement by Integration. In: Autonomous Intelligent Systems: Agents and Data Mining, Gorodetsky, V., J. Liu and V.A. Skormin (Eds.). Springer, New York, pp: 50-61.