International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34302-34306

A data mining approach in cloud for secure, scalable and efficient retrieval of data

S. Artheeswari, Dr.RM. Chandrasekaran
Research Scholar, Professor (CSE) / Director DDE,
Department of Computer Science, Annamalai University, Chidambaram.
artheeswariphd@gmail.com, aurmc@hotmail.com

Abstract- Cloud computing is a competitive alternative to general distributed data sharing scheme to share their user data with other, and perform their cooperative tasks together. Much of the data stored in clouds is highly sensitive, so it is necessary to prevent from attackers or unauthorized users. To maintain data privacy and efficient access, we proposed data mining scheme in cloud to provide better service to users. In this paper, we distribute data in decentralized cloud server and genetic algorithm for efficient retrieval. We utilized Attribute Based Encryption Scheme to enhance the security for our sensitive data. Thus our proposed security and access control scheme in cloud using data mining approach improves reliability and efficiency.

Keywords-*Cloud Computing, Data Distribution, Attribute Based Encryption Scheme (ASBE), Security.*

1. Introduction

The cloud platform is mainly used for storing and sharing the data between any two users or among the group. The cloud service provider (CSP) renders the software services, unlimited infrastructure that leads to maximum storage space. The users can access or share data exclusively from the cloud resources. The data in the cloud may have the confidential information that should not leak to the outsiders [1]. Considering the increasing number of users in the cloud, the security and privacy of the shared data on cloud becomes the main issue. Hence rendering the security in storing the data, confidentiality to the sensitive data and efficient retrieval of the data by the authorized user makes the cloud storage more elegance and effective [2].

Many cryptographic methods are available that encrypts the data, store it on the cloud and provide the decryption key when the authorized user request to access the data [3]. Traditional public key infrastructure has a data encryption process in which the data owner requires all the data users' public key to encrypt the data and results in overhead during the storage process which comprises of different public keys for the same data. Usually the information retrieval techniquefor example web based indexing are used to search and access the information from the web. The information retrieved should be in the reliable and flexible to get the optimal results and the cloud servers are susceptible to failures and attacks [4]. So it is necessary to implement the efficient storing and retrieval mechanism for the data.

However the provision of access control to the authorized user in a precise manner, achieving scalability of the storage system and strong key management process is not really fixed well in the previous cryptographic methods. The data in the cloud is stored in a centralized manner which is easier for the hacker to find the location of the encrypted data and a single point of failure disturb the whole data access process. In this paper, the Attribute Based Encryption Scheme is flexible and challenging approach to encrypt the data before outsourcing. Once the data is encrypted, it separated and stored in the cloud in a decentralized manner. The data are optimized using the genetic algorithm and the user on the other end retrieve the data using the decryption process of the ABE Scheme.

2. Related Work

The cloud data storage are mostly depends on the trusted third party and the access control and the data security becomes the major issue. The paper [5] discusses these issues and finds the solutions. The access control policy implementation based on the data qualities is processed and the fine grained access control is made on the untrusted cloud users and does not reveal the data. The Attribute Based Encryption based on the Key Policy is used for the secure and efficient data storage.

Many organizations like Personal Health Record are maintaining their patients' details in the centralized server and that should be having high security and the proper privacy. The users who accessing the details of the patients, should be assigned with the access rights based on their priority. All the data are updated in the third party cloud data server. Therefore the author suggested a new framework [6] that controls the data and the access mechanism in a semi trusted servers. In order to make it more secure, the ABE is used along with the option of multiple data owners called Multi Authority ABE.

The sensitive data should be encrypted before being saved in the centralized cloud environment. Traditional methods are available for encrypting and saving the data in the cloud platform and retrieve easily through the keywords. But these methods are having some drawbacks such as minimum tolerance and inconsistency in searching the keyword based data and possibility of frequent errors. The paper [7] suggested the fuzzy keyword search which increase the usability of the system by the way of search results. It gives the keyword based search results and also gives the nearest keyword matching data if exact match is not found. This technique considerably reduces the overheads in searching process and the storage process.

Data security and sharing in cloud storage brings many issues. The data sharing in cloud based on the dynamic

group key should have three features such as Key authentication, Key freshness and the Key confidentiality. Hence the paper [8] introduces the protocol for dynamic groupkey based on the Key Generation Center used to assign the key to the trusted members and cancel the permission the untrusted members. The author also introduces the short group signature scheme for data access control and the verification for cancelling the untrusted member data access.

Cloud renders large amount of storage for the cloud users and that should be more flexible and reliable for the users to save data and retrieve it without any difficulties. Though the cloud is gaining popularity because of its attraction and usability, it is still experiencing the issue in security and the privacy of data in cloud. The paper [9] intended the work of encryption algorithm to resolve this privacy and security problems.

Cloud computing in traditional way starts using the software and the local databases which does not give proper result in securing the data and cant able to store the data in reliable and flexible manner. Hence it moves the traditional way storing the data to large data centers where also the security challenges arise and leads to the two important and flexible parts composing a scheme [10]. The first part consists of homomorphic token and the second one is verification of erasure code data. This mechanism attains the data error localization and the accurate storage data.

3. Safe transmission of data in cloud

The organization or the institution handling big data over a distributed environment usually store the confidential data in the cloud platform. Since the data may contain some confidential information, the privacy and the security becomes a common issue. Therefore, many encryption methods are available to transform the data into cipher text for security and provide access controls to the authorized user for the preventing privacy of the data. The following sub-sections describe the encryption proposed encryption scheme, storage of data in the cloud platform and the retrieval of data from the cloud.

a) Encryption and Access Control

The data is encrypted using the Attribute Based Encryption Scheme (ABES) which is permits the user to access the data only if the user attributes satisfy the attributes of the cipher text. The data owner is the one who create and encrypt the original data. The users are those who attempt to access the data. Usually the encrypted data stored in the centralized server in the cloud and the users are attempting to access the data. Since there is possibility of finding the data location in the cloud and accessing it, the proposed system distributes the encrypted data over the cloud platform in a decentralized manner. The retrieval of encrypted data from different location of the cloud is possible using the genetic algorithm and their operations. The Attribute Based Encryption Scheme (ABES) is describes in the following sub-section.

Attribute Based Encryption Scheme (ABES)

The goal of the proposed system is to provide the secure access rights to the users and the owners for privacy preservation and the implement an efficient encryption scheme for providing the data security. Both the processes are integrated and works together like a mash-up application. Users who want to get access rights are supposed to submit their identity information to the data owner and get their own secret keys based on their attributes. This key is used to get the access rights of the particular data.User access rights are given based on the user privilege. Some users are given permission to do desired changes in the data or some may be allowed to read the desired data alone. In order to give the privacy preservation to the data, these access rights are given to the users (receivers).

In order to provide the access control, the Attribute base Access Control (ABAC) is used where the users are given an attribute or a set attributes and the access policy is attached with the data. This implies that the user those who are having the same attributes can able to access the day by satisfying access policies. The attributes used in the ABAC may be the data owner profile. It may be the years of experience, his/her birthday date or any of his/her profile details. For instance, in a college database, the faculties can able to access the particular class students mark list only if the faculty is handling any one subject for the particular class.Attribute Based Encryption Scheme (ABES) is known as primitive cryptographic method which is used in all ABAC tasks. Generally, the ABE has the two major classes namely Key Policy- Attribute Based Encryption (KP-ABE) and the Cipher text Policy-Attribute Based Encryption (CP-ABE). While using the KP-ABE [11], the data will be encrypted using a secret key and the with the access policy and has one or more attributes. The receiver on the other side can able to decrypt the data only if the attributes of the cipher text is matched with the attributes of the sender.In the CP-ABE, as in the KP-ABE, the CP-ABE is also use the same procedure with the major difference that the access policy in the receiver side is in the form of tree and the attributes are referred as the leaves of the tree.

These approaches use the Key Distribution Center (KDC) where in the attribute authority is synchronized with the data owner and intakes the attributes and the secret key from the encryption process and distribute them when required conditions are satisfied. The following steps are included in the Cipher text Policy-Attribute Based Encryption (CP-ABE) for the secure and access control data storage and retrieval process on cloud [12].

a)Global Setup (δ): This step takes the security parameter δ as input and brings the global parameter GP as output.

b) $\boldsymbol{Authority\ Setup\ (GP)}$***:*** The authority executes this step up by taking the Global Parameter (GP) as input and delivers the public key (P_k) and the secret key (S_k) of the authority as output.

c) $\boldsymbol{Encrypt\ (M,(H,\omega),GP,\{P_k\})}$***:*** The encryption process utilizes the plaintext M, global parameter GP, access matrix (H,ω) and the public key of the authority and finally brings the cipher text C.

d) $Key\ Gen\ (GID, I, GP, S_k)$ **:** the identity GID of the authority, the attribute i that belongs to the authority, the global parameter GP and the secret key which is generated for the authority are all taken as input for the key generation phase and produces the identity pair which consists of GID for the particular attribute and the key K_i.

e) $Decrypt\ (CT, \{K_i, GID\}, GP)$**:** the decryption process as usual takes the cipher text C, Global parameter GP and the identity pair $\{K_i, GID\}$ produced in the key generation pair as input and the output plaintext M is retrieved only if the attribute i is matched with the access matrix (H, ω) of the cipher text. If the attribute and the access matrix are not matched with each other, then the decryption process fails.

After the usual encryption process is done, the encrypted data from the data owner stored in the centralized server in the cloud platform. Since the encrypted data is stored in the same place, there is a possibility of hacking the data using some hacking techniques.

b) Search and Retrieval Process

Genetic Search Algorithm

Hence the proposed system makes use of the decentralized server in the cloud platform. Once the original message (Plaintext) is encrypted using the Attribute Based Encryption Scheme (ABES), the encrypted data is divided and stored in the cloud platform having decentralized server. A set of encrypted cipher text in which each cipher text in the cipher text set are having the individual attributes for each. For instances, considera school maintains record status for each class. The class teacher is the data owner of the record and each subject has its own attributes. The teachers of the particular subject have the attribute of the particular subject and the matches with the attribute of the cipher text. If both the attributes are matching, then the data owner provide him the secret key for that particular subject secret key or else it fails.

Once the data is divided into portions, they stored in the cloud platform in a decentralized way. This has the advantage of avoiding the location detection of the encrypted message. If the receiver wants the message for accessing, then the distributed encrypted message is gathered in a place and then sends for the decryption process. Since the cloud decentralized environment may have the data other than the ABES encrypted data, it is difficult to assemble the ABES encrypted data for decryption process.The genetic algorithm is used in this situation to gather the encrypted data in the cloud environment. It is the most efficient and elegance search technique and is the appropriate method for the data retrieval process.

As depicted in the figure 1, the genetic algorithm is a heuristic technique spread its chromosomes over the decentralized cloud platform [13]. Each chromosome represents the small portion of data. Then the fitness value is calculated for chromosomes that represents the small portion of data. The fitness value of each chromosome is computed by matching the attribute value of the data portion and the attribute value of the original data. The data portions which are having the relevant attribute value of the original plaintext are gathered and form the next generation chromosomes. All these data portions are combined in the format and check whether the cipher text attribute value and the original plaintext attribute value are exactly matching each other. If both the attributes are matched, then this means that the encrypted message is retrieved for the decentralized cloud computing platform. If not, then the genetic algorithm is continued with the newly generated chromosomes to retrieve the exact encrypted data. The above described genetic algorithm for efficient data retrieval from the decentralized cloud environment is shown in the figure 1.

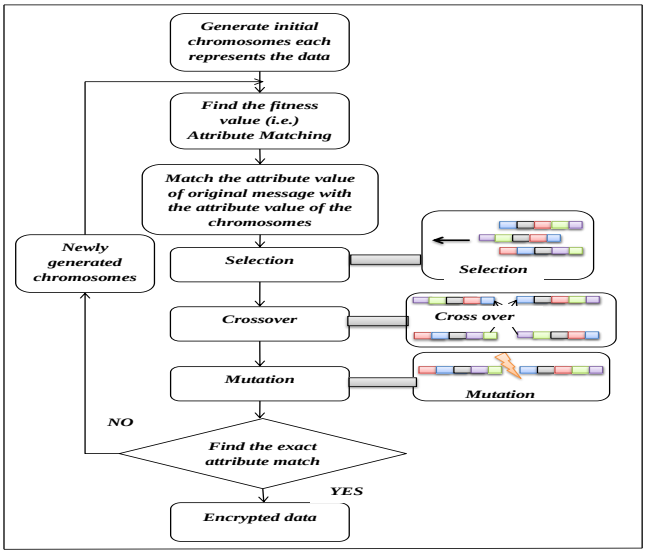

Fig 1 Genetic Data Retrieval

Once the encrypted data is retrieved, the original data attribute I is matched with the receiver side attribute. If both are same, then the decryption process is done using the parameters such as the cipher text C, Global parameter GP and the identity pair $\{K_i, GID\}$ produced in the key generation pair step. Finally the output will be the original plaintext message M.

4. Performance Analysis

The proposed system aims to retrieve the encrypted data from the cloud storage in secure and robust manner. Therefore it utilizes two algorithms and one approach to store and retrieve the data over the cloud platform in a very secure, scalable and efficient manner. Initially the Attribute Based Encryption Scheme (ABES) is used to encrypt the data based on the attribute of the data or the data owner. Then the encrypted data is stored in the cloud platform using decentralized approach. Final the genetic algorithm is used to gather the distributed data in the cloud using its fast searching behaviour. The gathered data is decrypted by the receiver after verifying the respective attributes.

Many searching algorithms are present to search the distributed data on the cloud. The existing two algorithms namely content based retrieval search and knowledge based retrieval search are compared with the genetic algorithm and proves that the genetic algorithm perform fast search and retrieval process against data in minimum time on the cloud as shown in the figure 2.

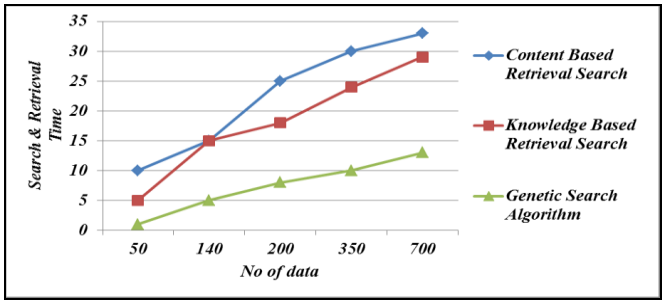

Fig 2 comparison of Genetic Search Algorithm with Existing Search Algorithms

There are many combination of encryption and retrieval techniques are present to transfer the data in a secure manner. Here the comparison of those combinations of techniques and the proposed system techniques are depicted in the figure 3.

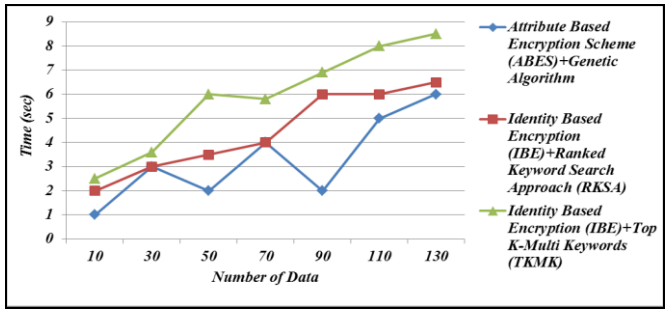

Fig 3 Comparisons of Proposed System Techniques with the Existing Techniques

The figure shows that the combination of ABES and the Genetic algorithm utilizes less time to store and retrieve the data against increasing number of data.

The figure 4 shows that the efficiency of the same existing techniques being compared with the proposed system techniques. In the result, the proposed techniques combination brings out better efficiency than the existing techniques.

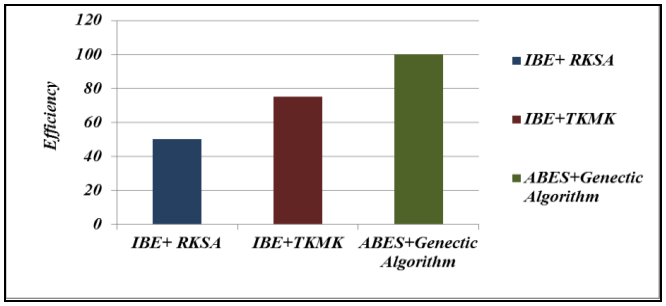

Fig 4 Efficiency of the Existing and Proposed System

5. Conclusion:

In this paper, the data is encrypted using the Attribute Based Encryption Scheme (ABES) based on the descriptive attributes. These encrypted data is stored in the cloud storage platform for easy sharing. In order to improve the security and robustness of the encrypted data, the decentralized approach is used. The encrypted data is stored in the decentralized format (distributed) over the cloud environment. Then the genetic algorithm which is a fast searching technique used to search the distributed encrypted data by using the attribute value of the data. After gathering and combining the encrypted data, the second part of the ABES is used to decrypt the data by matching the attributes of the original plaintext and the attributes of the encrypted text. The genetic algorithm for searching the distributed data is compared with the two existing algorithms and proves that the genetic algorithm shows better efficiency than those algorithms. The existing techniques used for the combination of encryption and searching the distributed data are compared with the proposed combination of techniques. Finally the proposed brings more efficiency and utilizes minimum time to process the data.

References:

[1] Yanjiang Yang and Youcheng Zhang, "A Generic Scheme for Secure Data Sharing in CLoud", Proceedings of 40th International Conference on Paraleel Processing Workshops (ICPPW), pp. 145-153, IEEE, ISSN: 1530-2016, September 2011.

[2] DananThilakanathan, Shiping Chen, Surya Nepal, Rafael Calvo and Leila Alem, "A Platform for Secure Monitoring and Sharing of Generic Health Data in Cloud", Journal of Future Generation Computer Systems, Volume 35, pp-102-113, June 2014.

[3] SharmilaRajasudhan and Nallusamy, "A Study on Cryptography Methods in Cloud Storage", International Journal of Communication and Computer Technologies Volume 02, Issue 02, ISSN: 2278-9723, March 2014.

[4] Zhigang Zhou, Hongli Zhang, Xiaojiang Du and Panpan Li, "Prometheus: Privacy-Aware Data Retrieval on Hybrid Cloud", Proceedings of INFOCOM, IEEE, pp-2643-2651, ISSN: 0743-166x, April 2013.

[5] S.SeenuIropia and R.Vijayalakshmi, "Decentralized Access Control of Data Stored in Cloud Using Key Policy Attribute Based Encryption", International Journal of Inventions in Computer Science and Engineering, Volume 1 Issue 2, ISSN (Online): 2348 – 3539, 2014.

[6] M. Vijayapriya and A. Malathi, "On Demand Security for Personal Health Record in Cloud Computing Using Encryption and Decryption Cryptography", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 9,ISSN: 2277 128X, pp.1083-1087, September 2013.

[7] Jin Li, Qian Wang, Cong Wang, Ning Cao, KuiRenand Wenjing Lou, "Fuzzy Keyword Search over Encrypted Data in Cloud Computing", Proceedings of IEEE INFOCOM, pp-1-5, ISSN: 0743-166x, March 2010.

[8] Dharani.R and M.Narmatha, "Secured Data Sharing with Traceability in Cloud Environment", International

Journal of Inventions in Computer Science and Engineering, Volume 1 Issue 8, ISSN (Online): 2348 – 3539,pp.1-9, September 2014.

[9] L. Arockiam and S. Monikandan, "Data Security and Privacy in Cloud Storage usingHybrid Symmetric Encryption Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, ISSN (Print): 2319-5940, pp. 3064-3070, August 2013.

[10] Cong Wang, Qian Wang, KuiRen and Wenjing Lou, "Ensuring Data Storage Security in Cloud Computing", 17thInternatioanl Work shop on Quality of Service, IEEE, pp-1-9, ISSN: 1548-615x, July 2009.

[11] M.Saranya and R.Vasuki, "Improving Data Security in KP-ABE with Third Party Auditing", International Journal of Inventions in Computer Science and Engineering, Volume 2 Issue 2, ISSN (Online): 2348 – 3539, pp-28-39, Feb 2015.

[12] S. Sankareswari and S. Hemanth, "Attribute Based Encryption with Privacy Preserving using Asymmetric Key in Cloud Computing", International Journal of Computer Science and Information Technologies, Volume 5, Issue 5, ISSN: 0975-9646, pp.6792-6795, 2014.

[13] Anubha Jain, Swati V. Chande and PreetiTiwari, "Relevance of Genetic Algorithm Strategies in Query Optimization in Information Retrieval", International Journal of Computer Science and Information Technologies, Volume 5, Issue 4, ISSN: 0975-9646, pp. 5921-5927, 2014.

[14] S. Artheeswari and Dr. R.M. Chandrasekaran, "An analysis on attribute based solutions for adaptable and scalable security access control in cloud computing" in International Journal of Computer Communication and Information System. . Vol 6., No. 3, July-September 2014, PP 96-105.

Author Profile

Mrs. S. Artheeswari is working as Assistant Professor in Mailam Engineering College, Mailam, Tamilnadu. She has 8 years of experience in academic field. She completed her Bachelor of Technology(IT) in Madras University and Master of Engineering(CSE) in Anna University. Now, doing as a Research Scholar in Annamalai University in the field of Computer Science. Her area of Interest includes Cloud computing, Data Structures, Security and Database Management System. She also has life member for several association and society.

Dr. RM. Chandrasekaran is currently working as a Professor, Department of Computer Science & Engineering and also jointly as Director, Directorate of Distance Education. Annamalai University. He obtained his Bachelor of Engineering in Computer Science and Master of Engineering from Anna University and Master of Business Administration from Annamalai University. Completed his PhD in Computer Science from Annamalai University, Annamalainagar, India. He has 23 years of teaching experience and 5 years of experience in Research & Development. He also worked as a Registrar in Anna University, Trichy for 3 Years. He worked as software consultant in USA. Also, he has co-organized two Workshops and two conferences. His area of interest includes Computer Algorithms, Text Data Mining and Software Metrics. He published 10 papers in National Journal and 10 papers in International Journals. He also published many papers in national and international conferences