International Journal of Current Research in Multidisciplinary (IJCRM)
ISSN: 2456-0979
www.ijcrm.com Volume 1 Issue 1 ǁ May. 2016 ǁ PP 01-05

High Dimensional Data's sharing in Multi Users Cloud Environment using Big Data Classification Technique

Saravanakumar M[1], Balamurugan M[2] , Santhoshkumar C[3]
[1,2,3](Computer Science, The Kavery Engineering College / Anna University , India)

Abstract:
The Cloud Computing, becomes sensitive information are being increasingly centralized into the cloud. In support of the protection of data privacy, responsive data has to be encrypted previous to outsourcing, which makes effective data consumption a very difficult task. Although traditional searchable encryption schemes allow users to securely search over encrypted data throughout keywords, these technique support only Boolean search, without capturing any relevance of data files. This come up to suffers from two main drawbacks, Not using centralized data sharing mechanism, Also the perturbed data do not produce very accurate data mining results, who do not necessarily have pre-knowledge of the encrypted cloud data, to post process every retrieved file in order to find ones most matching their interest; On the other hand over, invariably recover all files contain the query keyword additional incur pointless network traffic, we propose to further extend the scheme is big data classification technique to consider the extensibility of the file locate and the multi-user cloud environments. On the way to this direction.Various initial results on the extensibility and the multiuser cloud environments. An extra interesting topic is to enlarge the very much scalable searchable encryption to enable capable search on huge practical databases. Here explore a AES based encryption technique where access key generation for user is based on access policies assigned to each user along with the attributes. The data stored in the cloud is encrypted using a key generated based on the access permissions assigned to the data and attributes of the owners who share their data with high security and integrity using policy based encryption technique.

Keywords- *Searchable encryption, Multi-keyword, Big data, Cloud computing.*

I. Introduction

The cloud computing paradigm is revolutionizing the organizations' way of operating their data particularly in the way they store, access and process data. As an emerging computing paradigm, cloud computing attracts many organizations to consider seriously regarding cloud potential in terms of its cost-efficiency, flexibility, and offload of administrative overhead. Most often, organizations delegate their computational operations in addition to their data to the cloud. Despite tremendous advantages that the cloud offers, privacy and security issues in the cloud are preventing companies to utilize those advantages. When data are highly sensitive, the data need to be encrypted before outsourcing to the cloud. However, when data are encrypted, irrespective of the underlying encryption scheme, performing any data mining tasks becomes very challenging without ever decrypting the data. There are other privacy concerns, demonstrated by the following example. Suppose an insurance company outsourced its encrypted customers database and relevant data mining tasks to a cloud. When an agent from the company wants to determine the risk level of a potential new customer, the agent can use a classification method to determine the risk level of the customer. First, the agent needs to generate a data record q for the customer containing certain personal information of the customer, e.g., credit score, age, marital status, etc. Then this record can be sent to the cloud, and the cloud will compute the class label for q. Nevertheless, since q contains sensitive information, to protect the customer's privacy, q should be encrypted before sending it to the cloud.

Existing work on privacy-preserving data mining (PPDM) (either perturbation or secure multi-party computation (SMC) based approach) cannot solve the DMED problem. Perturbed data do not possess semantic security, so data perturbation techniques cannot be used to encrypt highly sensitive data. Also the perturbed data

do not produce very accurate data mining results. Secure multi-party computation based approach assumes data are distributed and not encrypted at each participating party.
In addition, many intermediate computations are performed based on non-encrypted data. As a result, in this paper, we proposed novel methods to effectively solve the DMED problem assuming that the encrypted data are outsourced to a cloud. Specifically, we focus on the classification problem since it is one of the most common data mining tasks. Because each classification technique has their own advantage, to be concrete, this paper concentrates on executing the k-nearest neighbor classification method over encrypted data in the cloud computing environment

II. Literature survey

2.1 Privacy-Preserving and Outsourced Multi-User k-Means Clustering

Clustering is one of the commonly used tasks in various data mining applications. Briefly, clustering [1]–[3] is the unsupervised classification of data items (or feature vectors) into groups (or clusters) such that similar data items reside in the same group. It has immense importance in various fields, including information retrieval [4], machine learning [5], pattern recognition [6], image analysis [7], and text mining [8]. Some real-life applications related to clustering include categorizing results returned by a search engine in response to a user's query, grouping persons into categories based on their DNA information, etc.In general, if the data involved in clustering belongs to a single entity (hereafter referred to as a user), then it can be done in a trivial fashion. However, in some cases, multiple users, such as companies, governmental agencies, and health care organizations, each holding a dataset, may want to collaboratively perform clustering task on their combined data and share the clustering results. Due to privacy concerns, users may not be willing to share their data with the other users and thus the distributed clustering task1 should be done in a privacy-preserving manner. This problem, referred to as privacy-preserving distributed clustering (PPDC), can be best explained by the following example: • Consider two health agencies (e.g., the U.S. CDC and the public health agency of Canada) each holding a dataset containing the disease patterns and clinical outcomes of their patients. Since both the agencies have their own data collecting methods, suppose that they want to cluster their combined datasets and identify interesting clusters that would enable directions for better disease control mechanisms.However, due to government regulations and the sensitive nature of the data, they may not be willing to share their data with one another. Therefore, they have to collaboratively perform the clustering task on their joint datasets in a privacy-preserving manner. Once the clustering process is done, they can exchange necessary information (after proper sanitization) if needed. The existing PPDC methods (e.g., [9]–[12]) incur significant cost (computation, communication and storage) on the participating users and thus they are not suitable if the users do not have sufficient resources to perform the clustering task.

2.2Public-Key Encryption with Data Sharing in Dynamic Groups for Mobile Cloud Storage

Mobile cloud computing is referred as the combination of cloud computing and mobile networks to bring benefits for both mobile users and cloud computing providers. While once the data of mobile users is outsourced to the cloud, it is a formidable and challenging task for the data owners to realize both the data confi- dentiality and the utilization because it seems unachievable to search and retrieve the special contents on the data encrypted by traditional encryption schemes. To address this issue, we propose a searchable public-key encryption scheme for a group of users in mobile cloud storage. In our proposal, a dynamic asymmetric group key agreement protocol is utilized for data sharing among a body of mobile users and the technique of proxy re-signature is employed to update the searchable ciphertexts when the mobile users in the group varies. Through the security proof and performance evaluation, we demonstrate the new scheme is both secure and efficient, and hence it reaches the requirements of the users, network operators, as well as cloud computing providers in application

The system public key encryption with keyword search scheme that supporting data sharing among multiple mobile users in a dynamic group. As far as we know, our work is among the first few ones to achieve the privacy preserving keyword search on encrypted data in mobile cloud storage.We motivate the searchable public-key encryption with data sharing for dynamic groups in mobile cloud storage and describe the system model and security threats. Deriving from the group key agreement protocol and proxy re-encryption, we propose a searchable encryption scheme which provides data sharing, group dynamic and efficient cipher texts updating. 3. We prove the security and justify the performance of our scheme by analyzing the computation, communication and storage overhead.

2.3 Multi-Owner Data Sharing in Cloud Storage Using Policy Based Encryption

Cloud storages are generally hosted by third parties where data can be stored and hared. Cloud storage provides virtualized pools of storage and people buy or lease storage capacity from them. The security

of data is major problem when people use commercial cloud services to store their data. To avoid unauthorized access, data should be encrypted before outsourcing. Instead of attribute based encryption, role based policies can be generated and based on that policies encryption can be done. In the case of data with multiple owners, the access control, integrity and revocation are major issues.
All the owners must have same access policy and revocation should be done with the permission of all owners. Another major issue is key generation and management. Here we explore a policy based encryption technique where access key generation for user is based on access policies assigned to each user along with the attributes. The data stored in the cloud is encrypted using a key generated based on the access permissions assigned to the data and attributes of the owners who share their data with high security and integrity using policy based encryption technique. Cloud computing is a general term for anything that involves delivering hosted services, scalable services like data sharing, accessing etc., over the web on demand basis. It uses the web and central remote servers to maintain data and applications. Cloud computing allows consumers and businesses to use applications without installation and access their personal files at any computer with web access. This technology allows for much more efficient computing by centralizing storage, memory, processing and bandwidth. Cloud computing is broken down into three segments: "application" "storage" and "connectivity". Each segment serves a different purpose and offers different products for businesses and individuals around the world.

2.4An Efficient and Secure Protocol for Querying High-Dimensional Data in the Cloud Technical Report
Data-as-a-service (DaaS) is a cloud computing service that emerged as a viable option to businesses and individuals for outsourcing and sharing their collected data with other parties. Although the cloud computing paradigm provides great flexibility to consumers with respect to computation and storage capabilities, it imposes serious concerns about the confidentiality of the outsourced data as well as the privacy of the individuals referenced in the data. In this paper we formulate and address the problem of querying encrypted data in a cloud environment such that query processing is confidential and the result is differentially private. We propose a framework where the data provider uploads an encrypted index of her anonymized data to a DaaS service provider that is responsible for answering range count queries from authorized data miners for the purpose of data mining. To satisfy the confidentiality requirement, we leverage attribute based encryption to construct a secure kd-tree index over the differentially private data for fast access. We also utilize the exponential variant of the ElGamal cryptosystem to efficiently perform homomorphic operations on encrypted data. Experiments on real-life data demonstrate that our proposed framework can efficiently answer range queries and is scalable with increasing data size.

III. Methodology

Multi-user cloud environment cloud environment using Big Data Classification Technique. To develop the highly scalable searchable encryption. Secure overlay cloud storage system that achieves key encryption control is proposed. the proposed architecture uses some important security services including authentication, encryption and decryption. It associates outsourced files with file access policies, and assuredly deletes files to make them unrecoverable to anyone upon revocations of file access policies. To achieve such security goals of secure key to sharing file system of key process.

COMPENSATION

- To enable efficient search on large practical database
- Centralized security control in cloud server.
- Automate deployment in data sharing in cloud server.
- Time limit for data sharing in cloud server.
- Unauthorized user not hacking the data in cloud server.

OVERVIEW OF THE BIG DATA CLASSIFICATION

A classic domain like stock market data are constantly generating a large quantity of information such as bids, buys and puts, in every single seconds [2]. This information impact on different factors such as domestic and international news, government reports and natural disasters and so on, hence it is nearly impossible to have required and appropriate information to user over such a complex and voluminous data so it is crucial that such a data should be classified appropriately and presented to the user for his convenience and ease of access. Classification technique is used to solve the above challenges which classify the big data according to the format of the data that must be processed, the

type of analysis to be applied, the processing techniques at work, and the data sources for the data that the target system is required to acquire, load, process, analyze and store. To illustrate these techniques

I. FIGURES AND TABLES

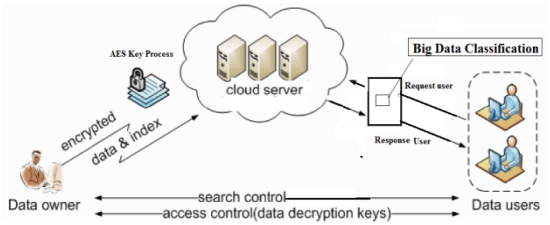

Fig 1.0, System Architecture

IV. Conclusion

We contain investigate on the using Big Data Classification Technique using multi-keyword search issue over encrypted cloud data, and proposed method. The BDCT include equally the significance score and the partiality factors of keywords to improve more specific explore and better user's experience, respectively. Furthermore, we have proposed the enhanced BDCT method supporting classified to improve efficiency of cloud data in data sharing. we design a secure data sharing with policies for dynamic groups of owners in a cloud storage system. During this, a user is bright to share data with others in the system as well as a group is also accomplished of storing and sharing their data. Moreover, this system supports efficient file revocation n and policy changing. More specially, efficient file revocation can be achieved throughout file policy revocation. So users cannot decrypt files stored in the cloud. Moreover, the storage overhead and the encryption computation cost are constant. Extensive analyses show that our proposed scheme satisfies the desired security requirements and guarantees efficiency as well.

REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen, and D. Peng, "An smdpbased service model for interdomain resource allocation in mobile cloud networks," IEEE Transactions on Vehicular Technology, vol. 61, no. 5, pp. 2222–2232, 2012.
[2] M. M. Mahmoud and X. Shen, "A cloud-based scheme for protecting source-location privacy against hotspot-locating attack in wireless sensor networks," IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 10, pp. 1805–1818, 2012.
[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo, "Exploiting geodistributed clouds for e-health monitoring system with minimum service delay and privacy preservation," IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 2, pp. 430–439, 2014.
[4] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong, and L. Zhang, "Privacypreserving data aggregation without secure channel: multivariate polynomial evaluation," in Proceedings of INFOCOM. IEEE, 2013, pp. 2634–2642.
[5] Y. Yang, H. Li, W. Liu, H. Yang, and M. Wen, "Secure dynamic searchable symmetric encryption with constant document update cost," in Proceedings of GLOBCOM. IEEE, 2014, to appear.
[6] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikeyword ranked search over encrypted cloud data," IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.
[7] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proceedings of S&P. IEEE, 2000, pp. 44–55.
[8] R. Li, Z. Xu, W. Kang, K. C. Yow, and C.-Z. Xu, "Efficient multikeyword ranked query over encrypted data in cloud computing," Future Generation Computer Systems, vol. 30, pp. 179–190, 2014.
[9] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen, "Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage," IEEE Transactions on Emerging Topics in Computing, 2014, DOI10.1109/TETC.2014.2371239.
[10] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proceedings of ICDCS. IEEE, 2010, pp. 253–262.
[11] A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill, "Order-preserving symmetric encryption," in Advances in Cryptology-EUROCRYPT. Springer, 2009, pp. 224–241.
[12] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Verifiable privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," IEEE Transactions on Parallel and Distributed Systems, vol. DOI: 10.1109/TPDS.2013.282, 2013.
[13] J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li, "Towards secure multikeyword top-k retrieval over encrypted cloud data," IEEE Transactions on Dependable and Secure Computing, vol. 10, no. 4, pp. 239–250, 2013.
[14] A. Arvanitis and G. Koutrika, "Towards preference-aware relational databases," in International Conference on Data Engineering (ICDE). IEEE, 2012, pp. 426–437.
[15] G. Koutrika, E. Pitoura, and K. Stefanidis, "Preference-based query personalization," in Advanced Query Processing. Springer, 2013, pp. 57–81.
[16] B. Zhang and F. Zhang, "An efficient public key encryption with conjunctive-subset keywords search," Journal of Network and Computer Applications, vol. 34, no. 1, pp. 262–267, 2011.

[17] D. Stinson, Cryptography: theory and practice. CRC press, 2006.
[18] H. Li, D. Liu, K. Jia, and X. Lin, "Achieving authorized and ranked multi-keyword search over encrypted cloud data," in Proceedings of ICC. IEEE, 2015, to appear.
[19] S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, and S. Mitra, "Zerber: r-confidential indexing for distributed documents," in Proceedings of the 11th international conference on Extending database technology: Advances in database technology. ACM, 2008, pp. 287– 298.
[20] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proceedings of SIGMOD International Conference on Management of data. ACM, 2009, pp.139–152.