INDEXING AND CLASSIFICATION OF ELEMENTS WITH ACCESS PRIORITIES OVER ENCRYPTED MULTI-CLOUD DATA

Antony Adhiban[1], Charlesmani[1] and Prem Jacob[2]
[1]Sathyabama University, Chennai, Tamilnadu, India
[2]Faculty of Computing, Sathyabama University, Chennai, Tamilnadu, India
E-Mail: gtadhib@gmail.com

ABSTRACT

In this paper, we address the problem of searching data in a group of encrypted data which is hosted by cloud. Cloud computing allows us to make data available online through the internet. Cloud computing is a concept where servers are connected to the internet and allowed for access to the public or private person using interconnected network. This, although a very good feature, does not provide privacy. There are risks of data being stolen or edited without the authorization of the owner. This in turn leads to security risks. If the data is encrypted using an encryption algorithm, it is hard to search the content for a specific data. The search capacity has to be compromised for the sake of security. This problem of searching encrypted data has been overcome by the usage of FMS. There are three references that we can take from already existing technological contributions. First, the usage of preference factor in cloud computing to give a neat UI experience for the customer or user who tries to search for a data in our search work. Second, the usage of logical operations inside the encrypted data in order improves speed and reduces size. Finally the concept of trapdoor has been taken from previous original contributions, which serves as the search methodology in a real-world dataset for ensuring privacy in search. This method is used to improve the overall accuracy and efficiency of the search and functional ability of the code.

Keywords: searchable encryption, multi-keyword, fine-grained, cloud computing.

INTRODUCTION

Cloud computing is a scalable and highly functioning framework. This framework consists of servers and access to the internet. These servers are connected through the internet and allowed for public data access. This allows the public to store and retrieve data from another system that is connected to the internet. There are various constraints that help and allow the user to navigate between the various entities that align themselves with the encrypted data. The searched items are suggestive: which means that, they are predatory in nature, taking what's needed and giving the desired output. If a user is to type out a particular search option - then it would be treated as only what is requested in the search would be represented on the screen, nothing more, and nothing less. This leads to a, To Whom It May Concern point of view, which enhances and provides added security. The pathways created have been considered from our original papers or contributions that have been taken. The most important factor of cloud computing is that it does not require any kind of permissions from a single user.

Cloud computing is the future of computing. In the near future, every single process, platform and storage can be provided as a service. This brings cloud computing as the centre of the universe of computing. The servers become the backbone of computing and internet naturally becomes the blood of the body of computing. If cloud computing can bring all the services and platforms to the framework, then there will be no requirement for having local storage or local processing power. Cloud computing in importance of business, it enables the environment of network by giving shared resource.it involves service to network. Software as a service (SaaS) implies in providing application for a business. The other kind here called as platform as a service (PaaS) which obtain custom application which can create business on their individuals own kind

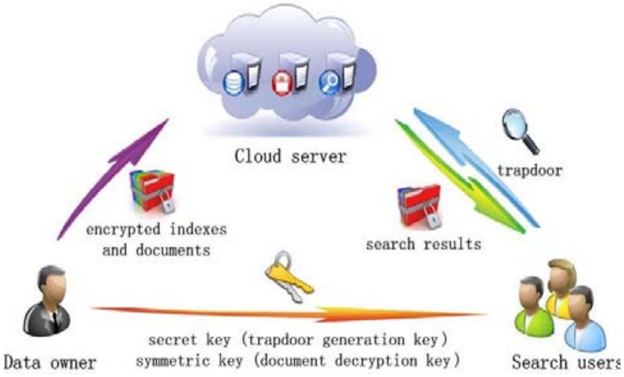
The only drawback of cloud computing lies in the connectivity or the bandwidth speed of the internet. This limitation can bottleneck the power of cloud computing and its uses. For any simple storage or platform service the bandwidth requirement is very high, therefore, maybe in the near future there are possibilities that the bandwidth can match the requirements for cloud services.

OWNER-Data upload: Anyone can upload data to the cloud and allow public access for the data he has uploaded. The public access refers to any user who requires data and searches using keyword. The data that is uploaded to the cloud requires privacy. This privacy can be achieved by encrypting the data using encryption methods. Keywords can be generated to improve the efficiency of search speed. The link between the keywords generated for the index and the corresponding encryption key for the data is created. Finally the search user can search using keywords. As a result the search user gets the decryption key for the data that he has requested.

SERVER-Cloud: The cloud is any server or system containing data that is connected to the internet and available for user access through the web. The cloud computing not only allows for data access but also service access and platform access. The cloud server acts as a bridge between the data downloader and data uploader. The data and its corresponding index is always stored in cloud. When someone with authorization from the administrator searches the cloud using trapdoor. The cloud in turn returns the result after apply a set of search operations.

USER-Search: There are three steps involved when a user requests for a search.

- The user gets two keys, data encryption key and data search key or the trapdoor key.
- After getting the both keys, the user is allowed to generate the portal which can allow access to the data
- The user can now get the documents and decrypt using the key provided.



Privacy and security requirements

The privacy model of this setup is said to be honest but curious, which means the privacy model always sticks to the protocol it is designated and the specification of the system is also checked every time it is implemented. Although the setup always follows the rules, it has a curiosity to change and analyze the data inside the index that is created to return keywords in order to learning new information. There are two methods which we are basing our privacy model on.

- Existing encryption model: Analyzing is only done in encrypted document and encrypted index that is uploaded by the uploading user.
- Existing Acknowledgement model: The cloud can have much higher power considering the encryption of the data, which uses an acknowledgement relationship of the dataset.

The mutual understanding between the user and administrators is the key to working of the search encryption. The administrator can provide user access or remove them. The data owner’s trust of the administrator also matters in this case. The person who uploads must be in confidence with the administrator. The whole concept is based on the trust between the user, administrator and the data owner.

The problem of the user needing to request permission every time from the administrator is a very frustrating process for continuous or simultaneous data users. This can be prevented by providing the user with central access using a login into the webpage.

The Security requirements can be derived from our privacy model in terms as the following content.

- The data that has been uploaded by the user has to be safely stored in the cloud using a highly secure encryption algorithm such as AES, DES etc.
- Due to the encryption of data, the data might not be available for search.
- The cloud not must be able to link any details that can be linked between the keyword index and the trapdoor. If the cloud is able to link these details then there might be security risks or privacy leaks.
- The regression and classification of the keywords allows for separating the gap between the index and data available in the cloud.
- The files stored in the cloud is searched or requested by many users over a long period of time. This might allow users to get an idea of what the actual data present in each position of the index.
- Therefore, to prevent assumption or security risks, the trapdoors have to be highly encrypted and often refreshed in order to prevent any data leak
- As in aspects to sub-sets of data, leakages does not happen in data sub-sets even if the user has complete access to the database.

Algorithm and concept

a) The k nearest neighbor algorithm is the simplest and one of the first finds the Euclidean distance between two elements inside a set consisting of classified data. There are three variables, first, a vector that is used to split the text, second and third are keys that are used to encrypt the data stored as the first element.

b) The below given formula is used to calculate the relevance between the search keyword and the possibility of the value present in the cloud.

c) Frequency is the number of times the keyword can be found inside the dataset.

d) Prevalence is the factor that it may either be present or not present inside the document.

Performance

The performance for every single aspect of this setup can be compared and revision made. The encryption used in this setup is AES. AES is used because it is the highest level of security available. Most encryptions used in cloud computing has been DES. But DES does not provide enough privacy. DES is broken or can be cracked over a long hacking exposure. The AES has not yet been cracked, but will be cracked over the period using the supercomputers that are available right now. The future of encryption lies in homomorphism. On a different point of view, the time taken to build the encryption algorithm is the amount of security for its cryptographic nature.

The network that binds platforms like computers, mobiles and other hardware devices is software defined networks. These types of network are known for its most defining features like unpredictability and self-healing effect on its architecture. The SDN is cost-effective, has a scalable bandwidth and has high performance range. This is the future of any business or art field.

The k-NN computation is a simple machine learning algorithm that is a type of lazy or instance learning. It has two purposes, which are classification and regression. Classification is the process of finding similarities and dissimilarities and arranging the elements in proper order. Regression is the process of repeating until finding the optimal solution. The accuracy of the k-NN computation can be found using matching matrix or a much effective result can be found with ration test.

Regression testing allows us to find the optimal solution for secure data encryption which comes as a final solution to AES. The k-NN computation allows for finding the next element in a database.

The file storage in the database can be upgraded by changing the data type to long blob function and updating or saving the database.

The SQL server management is the most widely used type of database tool that is highly scalable, efficient and robust. This tool is used to create a database and connect it to a live working code in order to store data. Cloud computing is entirely based on databases. These databases can be created and manages using SQL server manager. The most important aspect in this paper for databases is for the databases to be encrypted and secure. If someone can get access to any of the database files then there is a high risk of the entire system getting hacked. Therefore for this purpose we are using k-NN computation, which protects the databases like a pseudo-code.

The most important aspect of databases with security is using data sub-sets. These subsets do not link directly with any elements inside the code. Instead it becomes a sub-code that can be connected with the original elements of the database. For example, there are twenty elements that belong to the same category, then the database uses k-NN computation to automatically create a sub-set that contains all elements under a different name, in order to improve security and privacy.

REFERENCES

[1] H. Liang, L. X. Cai, D. Huang, X. Shen and D. Peng. 2012. An smdpbased service model for interdomain resource allocation in mobile cloud networks. IEEE Transactions on Vehicular Technology. 61(5): 2222-2232.

[2] Amit Sahai Ciphertext-Policy Attribute-Based Encryption in Proc. IEEE Symposium on Security and Privacy, Oakland, USA, 2007, pp. 321-334.

[3] Q. Shen, X. Liang, X. Shen, X. Lin, and H. Luo. 2014. Exploiting geodistributed clouds for e-health monitoring system with minimum service delay and privacy preservation. IEEE Journal of Biomedical and Health Informatics. 18(2): 430-439.

[4] T. Jung, X. Mao, X. Li, S.-J. Tang, W. Gong and L. Zhang. 2013. Privacy preserving data aggregation without secure channel: multivariate polynomial evaluation. in Proceedings of INFOCOM. IEEE. pp. 2634-2642.

[5] Y. Yang, H. Li, W. Liu, H. Yang and M. Wen. 2014. Secure dynamic searchable symmetric encryption with constant document update cost. in Proceedings of GLOBCOM. IEEE.

[6] Vigneshwari S and Aramudhan. M. 2015. Personalized cross ontological framework for secured document retrieval in the cloud. National Academy Science Letters-India. 38(5): 421-424. DOI 10.1007/s40009-015-0391-3, scopus, web of science.

[7] https://support.google.com/websearch/answer/173733?hl=en.

[8] D. X. Song, D. Wagner and A. Perrig. 2000. Practical techniques for searches on encrypted data. in Proceedings of S and P. IEEE. pp. 44-55.

[9] R. Li, Z. Xu, W. Kang, K. C. Yow and C.-Z. Xu. 2014. Efficient multikeyword ranked query over encrypted data in cloud computing. Future Generation Computer Systems. 30: 179-190.

[10] H. Li, D. Liu, Y. Dai, T. H. Luan, and X. Shen. 2014. Enabling efficient multi-keyword ranked search over encrypted cloud data through blind storage. IEEE Transactions on Emerging Topics in Computing, DOI10.1109/TETC.2014.2371239.

[11] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou. 2010. Secure ranked keyword search over encrypted cloud data. in Proceedings of ICDCS. IEEE. pp. 253-262.

[12] A. Boldyreva, N. Chenette, Y. Lee, and A. Oneill. 2009. Order-preserving symmetric encryption. in Advances in Cryptology-EUROCRYPT. Springer. pp. 224-241.

[13] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li. 2013. 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.

[14] J. Yu, P. Lu, Y. Zhu, G. Xue, and M. Li. 2013. Towards secure multikeyword top-k retrieval over encrypted cloud data. IEEE Transactions on Dependable and Secure Computing. 10(4): 239-250.

[15] A. Arvanitis and G. Koutrika. 2012. Towards preference-aware relational databases. in International Conference on Data Engineering (ICDE). IEEE. pp. 426-437.

[16] G. Koutrika, E. Pitoura, and K. Stefanidis. 2013. Preference-based query personalization. in Advanced Query Processing. Springer. pp. 57-81.

[17] B. Zhang and F. Zhang. 2011. An efficient public key encryption with conjunctive-subset keywords search. Journal of Network and Computer Applications. 34(1): 262-267.

[18] D. Stinson. 2006. Cryptography: theory and practice. CRC press.

[19] H. Li, D. Liu, K. Jia and X. Lin. 2015. Achieving authorized and ranked multi-keyword search over encrypted cloud data. in Proceedings of ICC. IEEE.

[20] S. Zerr, E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett, and S. Mitra. 2008. Zerber: r-confidential indexing for distributed documents. in Proceedings of the 11th international conference on Extending database technology: Advances in database technology. ACM. pp. 287-298.

[21] W. K. Wong, D. W.-l. Cheung, B. Kao, and N. Mamoulis. 2009. Secure knn computation on encrypted databases. in Proceedings of SIGMOD International Conference on Management of data. ACM. pp. 139-152.

[22] J. Zobel and A. Moffat. 1998. Exploring the similarity space. in ACM SIGIR Forum. 32(1). ACM, pp. 18-34.

[23] N. Ferguson, R. Schroeppel and D. Whiting. 2001. A simple algebraic representation of rijndael. in Selected Areas in Cryptography. Springer. pp. 103-111.

[24] P. Asha, Dr.T. Jebarajan. 2014. Improved parallel pattern growth data mining algorithm. in the International Review on Computers and Software, ISSN 1828-6003, 9(1): 80-87.

[25] P. Golle, J. Staddon, and B. Waters. 2004. Secure conjunctive keyword search over encrypted data. in Applied Cryptography and Network Security. Springer. pp. 31-45.

[26] D. Boneh and B. Waters. 2007. Conjunctive, subset, and range queries on encrypted data. in Theory of cryptography. Springer. pp. 535-554.

[27] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. 2004. Public key encryption with keyword search. in Advances in Cryptology-Euro crypt. Springer. pp. 506-522.

[28] Y. Hwang and P. Lee. 2007. Public key encryption with conjunctive keyword search and its extension to a multi-user system. in Proceeding of Pairing. Springer. pp. 2-22.

[29] Q. Liu, C. C. Tan, J. Wu, and G. Wang. 2012. Efficient information retrieval for ranked queries in cost-effective cloud environments. in Proceedings of INFOCOM. IEEE. pp. 2581-2585.