SECURITY ANALYSIS AND PERFORMANCE EVALUATION ON PRODUCT INFORMATION BASED ON THE INDEPENDENT SECRET KEYS PRIVACY PRESERVING DATASEARCH SCHEME USING RANKING ALGORITHM

Kasina Uday Shankar[1], Vanithakakollu[2],K. Yasudha[3]
[1]PG Student, [(2,3)]Assistant Professor
Department of Computer Science, GIS, GITAM(Deemed to be University)

Abstract:

Cloud computing is a shows potential in recent decades Information technology that can organize a large amount of ability in an efficient and flexible manner. More and more frequent companies plan to go their local data management systems to the cloud and store and manage their product information on cloud servers. An additional heated discussion is how to preserve the protection of the commercially confidential data, while maintaining the ability to search the data .therefore In this paper, a privacy-preserving data search scheme is planned, that can support both the identifier based and feature-based product search. Uniquely two novel index trees are constructed and encrypted, that can be search without knowing the plaintext data. Analysis and simulation results demonstrate the security and effectiveness of our scheme.
Keywords: Cloud computing, Privacy preserving, Information security, Product information retrieval.

1. Introduction

Economic growth is very slow in world wide, there is a need to change the environment using e commerce technology through the internet the traditional technology will improve very expansion. All aspects of social and economic activities, thereby driving the development of enterprise-level ecommerce, both in scope and in depth, and facilitating the transformation and upgrading of enterprises. The Monitoring Report on the Data of China's Ecommerce environment , the volume of ecommerce transactions in the world reached approximately, e-retail sales accounted for 14.1%. The rapidly rising number of cyber-transactions has spawned ecommerce big data. As increasingly numerous data files are being stored locally in enterprises, the pressure on local data storage systems greatly increases. Local hardware failures lead to great damage or loss of data, which greatly affects the daily operations of the enterprises. Fortunately, cloud storage techniques came into being under such circumstances. Cloud computing can collect and organize a large number of different types of storage devices by means of various functions, such as cluster applications, network technology and distributed file systems. There have already been a number of typical cloud service products at home and abroad, such as Amazon Web Services , Microsoft Azure , i-Cloud , and App Engine As large amounts of data are outsourced to cloud storage servers, With the growth of the company, product information also increases greatly. To improve the stability and reliability of a data storage system, an intuitive scheme is moving the local data management system to the cloud. Cloud computing is widely treated as a promising information technique (IT) infrastructure because of its powerful functionalities. The contributions are as summarized as A product information outsourcing and searching system model including the data owner, cloud server and data users is designed. Then index structures supporting efficient product retrieval are constructed. Moreover, corresponding search algorithms are also proposed. The author integrate the secure ranking algorithm into our scheme to guarantee the security of the outsourced data while maintaining the search ability. the related work of privacy-preserving data search schemes organized , Next the data search system model and preliminary techniques are discussed and the encrypted product information retrieval scheme in detail, and the evaluation of the proposed scheme is provided .Finally, the study's conclusions are presented.

2. Related work

Cloud storage services have quite a lot of advantages, such as ease of use and cost economy, and they are broadly used in many fields. However, a number of challenges are associated with them. With the increasing popularity of cloud storage, security issues have turn into an important factor restrict its growth. In recent existence, data leakage accidents have frequently occurred in such companies as Microsoft, Google and Amazon, and these incidents have exacerbated

users fears. To counter the information leakage, data owners and enterprises typically outsource the encrypted business data, rather than the plaintext data, to cloud storage servers. In general, the outsourced data can be divided into three types. The first type is the open-resource-type data, which do not need to be hidden from the cloud server, such as the basic information of the enterprise and the parameters of products. The second type is the private data, which need to be encrypted but are only access and decrypted by the data provider. This type encrypted independently, and the users need to scan the entire manuscript to search for a certain keyword. as a result, this method has an extremely high searching complexity. Its formally built the security definitions for symmetric searchable encryption, and a scheme based on a Bloom filter was designed. The security definitions are extended in Due to the lack of a rank mechanism for the returned results, the data users need to take a long time to screen the returned results, which is unacceptable in general. Thus, many single keyword ranked search schemes have been proposed Though these schemes can return more accurate search results, they cannot satisfy users' requirements in most cases, considering that a single word cannot provide sufficient information to describe the users' interests.

Multiple keyword Boolean search schemes allow the data users to input a set of keywords to search the desired documents. Conjunctive keyword search schemes return the documents in which all the keywords specified by the search query appear; disjunctive keyword search schemes return all the documents that contains at least one keyword of interest. Predict keyword search schemes have been proposed to support both conjunctive and disjunctive search patterns. However, the returned results are still not sufficiently suitable to the users because the degrees of importance of the keywords are not considered in these schemes.

includes such data as internal confidential information, intellectual properties and patents. The third type is the private data that need to be encrypted but can also be shared with specific users or groups .This type includes internal shared data, hospital's division-wide case in order and information by some shared advanced users. A single keyword Boolean search is the simplest document retrieval method for encrypted files. The other researchers proposed the searchable encryption scheme in which each word in a document is

3. System Model:

To manage the product and collecting the product information the data manager is the responsible for all. In addition, the data manager needs to encrypt the product information file by a symmetric encryption technique before outsourcing the data to the cloud server. To improve the security of the files, each file is encrypted by a single secret key, and the keys of different files are independent. Furthermore, to improve the search efficiency, an

index structure is constructed for the outsourced data. At first, an identifier index structure is constructed based on the hash function and height-balanced binary search tree. Then, a feature vector tree is built for all the feature vectors of the product, and it is encrypted by the secure K-NN algorithm.

When a data user wants to search a set of chosen products, she needs to generate a trapdoor to describe her interest. Two types of the trapdoor can be provided, i.e., a set of

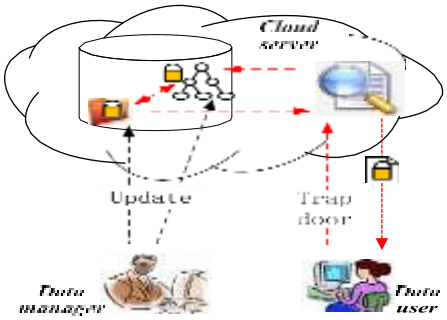


Figure 1. Retrieval system model Encrypted product information.

hash values of the desired product information files or a set of feature vectors. For the first type of trapdoor, a set of encrypted files with the same hash identifiers are returned, and for the second type trapdoor, the most relevant encrypted files are returned. The data user can obtain the plain text files by decrypting the returned files with the help of the symmetric secret keys. These secret keys are provided by the data manager.

The cloud server stores all the data uploaded by the data manager. When a data user needs to search the data in the cloud, she first generates a trapdoor, which is sent to the cloud server. A search engineer is employed by the cloud server to act as a bridge between the data users and the encrypted data. Though the cloud server cannot get the plaintexts of the data, it should be capable of sending the accurate search result of the trapdoor to the data users. Of course, the returned data are cipher text, and the data user needs to decrypt them by the symmetric secret keys which are provided by the data manager.

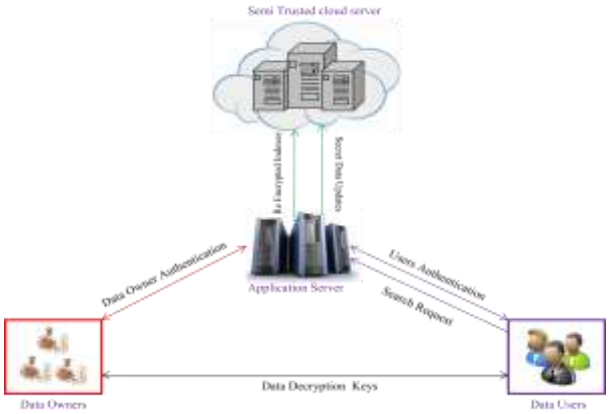


Figure 2. Cloud Architecture

To construct the ID-AVL tree, the first encrypt all the product identifiers based on a hash function, hash(). Next, each node in the ID-AVL tree contains a hash value of the product ID, and all of the hash values are organized based on an AVL tree

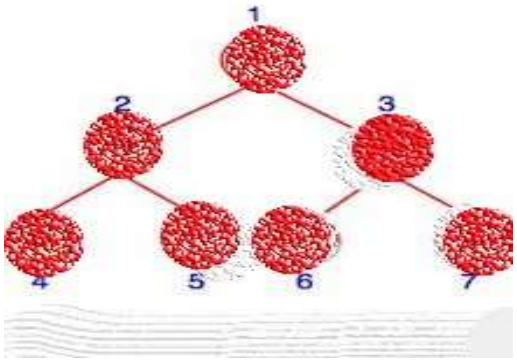


Figure 3: AVL Tree

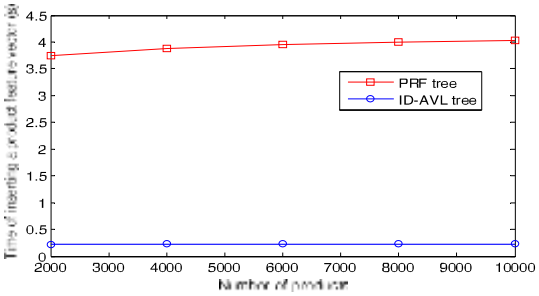


Figure: 4. Time of constructing the trapdoors.

Two important properties of AVL, which can help us to maintain the hash values, are presented as follows. First, the ID-AVL tree can be updated flexibly by inserting a node, deleting a node and modifying a node. Correspondingly, we can update the ID-AVL tree from time to time by changing the product information. then, the values of the left child nodes of a parent node are always smaller than that of the parent node, the values of the right child nodes of a parent node are always larger than that of the parent node. In theory, the time complexity of inserting, deleting and searching a node are all log(N), where N is the number of nodes in the tree. In this paper, we construct the ID-AVL tree based on the algorithm. Moreover, in certain cases, the data user may want to search the product based on the features. Initially, the data user needs to construct the feature vector of the product as discussed Then, the author need to design a depth-first search algorithm for the PRF tree, and that algorithm is presented in Depth first search Algorithm.

Consider u r, While u is not a leaf node, then Calculate all the relevance scores between the child nodes of u with V_Q ; u the most relevant child node; end while

Select the most relevant k document vectors in u by R Score(V_i, V_Q) and construct *RList*; s*tack.* push(*r*), while *Stack* is not empty u *Stack*.pop(), if the node u is not a leaf node, and if R Score $V_{u,max}$, V_Q^{Σ} >k thScore.

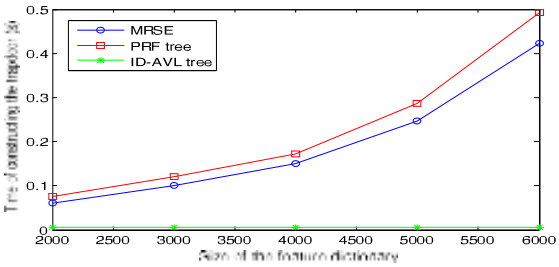
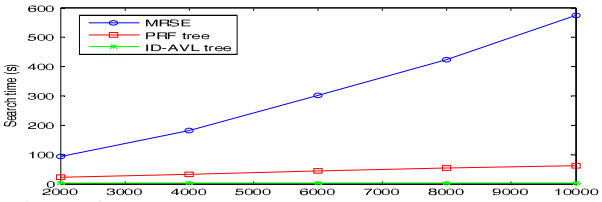


Figure: 5. Search Time With Different Number Of Products.

Figure: 6. Time use of insert a node



into the trees.

4. Methodology:

In security analysis of our scheme, the outsourced data includes the product information file, ID-AVL tree and PRF tree. The product information files are encrypted symmetrically based on the independent secret keys, and the cloud server does not have the tree cannot be recovered.

In the other way the Product Information search efficiency is to evaluate the search efficiency of our scheme.; First, we evaluate the construction time of the index structures of the product information. Specifically, the

author compare our scheme with the MRSE scheme. To decrease the bias of the data manager who is responsible for generating the vectors and the hash values, in this paper we employ the Enron Email Data Set to test our scheme. Specifically, the data set is employed to act as the product information files. Moreover, the vectors of the product are assumed to be extracted from the data set based on the TF-IDF model, and then the vectors are organized by the PRF tree. The increasing number of products, the construction time of PRF tree and the index structures in MRSE monotonously increase. This is reasonable considering that each product information file needs to be scanned for a time to get the feature vectors. The construction time of the PRF tree is slightly

increasing of the feature dictionary's size. This is reasonable considering that the size of the product feature vector is equal to the size of the feature dictionary. In addition, the time costs for the MRSE and PRF trees are similar to each other because the processes of generating the trapdoors are similar.

The search time of a trapdoor in the cloud server is presented in Fig. 5. It can be observed that the MRSE scheme consumes the most time to execute a search operation. Moreover, the search time increases monotonously with the increasing of the number of products. This increase can be explained by the fact that in MRSE, the feature vectors are stored in order, and they do not employ any index structure. In this case, the cloud server needs to

longer than that of the MRSE scheme, because the vectors need to be further inserted to the trees in the PRF tree. Apparently, the ID-AVL tree is considerably simpler, and the construction time can be ignored compared with the other two trees. To search the desired product information, the data user needs to first generate the trapdoor, which is sent to the cloud server. The times of constructing the trapdoors with the increasing of the size of the feature dictionary are presented in Fig. 3. The search requests based on the identifiers are independent of the feature dictionary, and hence, the time of constructing the trapdoors for the ID-AVL tree remains stable. However, the construction time of the trapdoors for the MRSE and PRF trees monotonously increase with the number of products,

scan all the product feature vectors to get the search result. The PRF tree organizes the vectors by a height-balanced tree, and most paths in the tree are pruned in the search process. As a consequence, the search efficiency is greatly improved. Finally, we can observe that the ID-AVL tree is the most efficient index structure, which can be explained by the fact that the ID-AVL tree is considerably simpler, and the search process is also very easy.

With the expanding growth of companies, more and more product information needs to be outsourced to the cloud server. Consequently, we need to update the index trees from time to time, and the update efficiency also affects the performance of our scheme significantly. As shown in Fig. 6, the update time of both the

PRF tree and the ID-AVL tree increases slightly with the increasing number of products, which is reasonable, considering that we need to search the trees to identify the proper location of the inserted node. In addition, updating the PRF tree consumes much more energy than that of updating the ID-AVL tree. This can be explained by the fact that the ID-AVL tree is much simpler than the PRF tree, and in theory only $\log(N)$ nodes need to be searched. Though quite many paths in the PRF tree are pruned in the search process, the number of the search paths is considerably larger than $\log(N)$ and more time is thus consumed in the PRF tree.

5. CONCLUSION

In this paper, we designed a secure and efficient product information retrieval scheme based on cloud computing. particularly, two index structures, including a hash value AVL tree and a product vector retrieval tree, are constructed, and they support an identifier based product search and feature vector based product search, respectively. Correspondingly, two search algorithms are designed to search the two trees. To protect the product information privacy, all the outsourced data are encrypted. The product information is symmetrically encrypted based on a set of independent secret keys, and the product vectors are encrypted based on the secure k-NN algorithm. Security analysis and simulation results illustrate the security and efficiency of the proposed scheme.

As the future work, we attempt to seamlessly integrate more index structures into our scheme to support more search patterns. Another difficult and promising challenge is further improving the search efficiency.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," *Communication of the ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[2] C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy preserving public auditing for secure cloud storage," *Computers, IEEE Transactions on*, vol. 62, no. 2, pp. 362–375, 2013.

[3] Dhananjaya, Balasubramani R, "A ccelerating Information Security inCloud Computing using a Novel Holomorphic Scheme", International Journal of Innovative Technology and Exploring Engineering (IJITEE)ISSN: 2278-3075,Volume-9Issue-1, November2019.

[4] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in *Proc. ACM CCS'06*, VA, USA, Oct. 2006, pp. 79–88.

[5] D. B. et al., "Public key encryption with keyword search secure against keyword guessing attacks without random oracle," *EUROCRYPT+*, vol. 43, pp. 506–522, 2004.

[6] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secure and Dependable Storage Services in Cloud Computing," IEEE Trans. Services Computing, vol+. 5, no. 2, pp. 220-232, Apr.-June 2012.

[7] T. Jung, X. Y. Li, Z. Wan, and M. Wan, "Privacy preserving cloud data access with multi-authorities," in *Proc. IEEE INFOCOM' 13*, Turin, Italy, Apr. 2013, pp. 2625–2633.

[8] Wei Zhang, Yaping Lin, Sheng Xiao, Jie Wu and Siwang Zhou, "Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing" January 2015.

[9] PL. Chithra, K. Sathya, "Color Channel CAPTCHA: A Secure Scheme for Cloud Environment", International Journal of Innovative Technology and Exploring Engineering (IJITEE)ISSN: 2278-3075,Volume-9Issue-1, November2019