Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

A RELIABLE AND DYNAMIC MULTIPLE KEYWORDS RANKED SEARCHING MECHANISMIN CLOUD

[1]Munnangi Sampath Kumari, [2]Y.Vijetha, [3]A.Chennakesavareddy, [4]Bathula Venkata Narendra
[1,2,3]Assistant Professor, [4]Student, Dept. of Computer Science Engineering, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

ABSTRACT
The emerging cloud computing paradigm, the data and corresponding complex management tasks can be outsourced to the cloud for the management flexibility and cost savings. Unfortunately, as the data could be sensitive, the direct data outsourcing would have the problem of privacy leakage. The encryption can be used, before the data outsourcing, with the concern that the operations can still be accomplished by the cloud. We consider the multi keyword similarity search over outsourced cloud data. In particular, with the consideration of the text data only, multiple keywords are specified by the user. The cloud returns the files containing more than a threshold number of input keywords or similar keywords, where the similarity here is defined according to the edit distance metric. We propose three solutions, where blind signature provides the user access privacy, and a novel use of Bloom filter's bit pattern provides the speedup of search task at the cloud side. Our final design to achieve the search is secure against insider threats and efficient in terms of the search time at the cloud side. Performance evaluation and analysis are used to demonstrate the practicality of our proposed solution**s.**
Keywords: capital costs, storage, processing, memory, network bandwidth, and virtual machines

I. INTRODUCTION

Cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). The name comes from the common use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation. Cloud computing consists of hardware and software resources made available on the Internet as managed third-party services. These services typically provide access to advanced software applications and high-end networks of server computers.

The cloud computing uses networks of large groups of servers typically running low-cost consumer PC technology with specialized connections to spread data-processing.

The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location-independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or data center). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

Rapid elasticity: Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage

can be managed, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

Achieve economies of scale – increase volume output or productivity with fewer people. Your cost per unit, project or product plummets. Reduce spending on technology infrastructure. Maintain easy access to your information with minimal upfront spending. Pay as you go (weekly, quarterly or yearly), based on demand. Globalize your workforce on the cheap. People worldwide can access the cloud, provided they have an Internet connection.

Streamline processes. Get more work done in less time with less people. Reduce capital costs. There's no need to spend big money on hardware, software or licensing fees. Improve accessibility. You have access anytime, anywhere, making your life so much easier! Monitor projects more effectively. Stay within budget and ahead of completion cycle times. Less personnel training is needed. It takes fewer people to do more work on a cloud, with a minimal learning curve on hardware and software issues. Minimize licensing new software. Stretch and grow without the need to buy expensive software licenses or programs. Improve flexibility. You can change direction without serious "people" or "financial" issues at stake.

II.RELATED WORK

The amount of data generated by individuals and enterprises is rapidly increasing. With the emerging cloud computing paradigm, the data and corresponding complex management tasks can be outsourced to the cloud for the management flexibility and cost savings. Unfortunately, as the data could be sensitive, the direct data outsourcing would have the problem of privacy leakage. The encryption can be used, before the data outsourcing, with the concern that the operations can still be accomplished by the cloud. We consider the multi-keyword similarity search over outsourced cloud data. In particular, with the consideration of the text data only, multiple keywords are specified by the user. The cloud returns the files containing more than a threshold number of input keywords or similar keywords, where the similarity here is defined according to the edit distance metric. We propose three solutions, where blind signature provides the user access privacy, and a novel use of Bloom filter's bit pattern provides the speedup of search task at the cloud side. Our final design to achieve the search is secure against insider threats and efficient in terms of the search time at the cloud side. Performance evaluation and analysis are used to demonstrate the practicality of our proposed solutions.

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

It is achieved by creating user-friendly screens for the data entry to handle large volume of data. The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.

When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow

Recently, privacy-preserving keyword search, such as secure ranked search, where only the files with better matching to the input keywords are returned, has been studied in the settings of single keyword and multi-keyword.

Nonetheless, in the setting of secure ranked search, only the number of keyword matches is concerned, and the similarity between the input keywords and the actual words in text is not taken into account.

On a different front, privacy assured similarity search, where the files containing exactly the same keyword or containing similar keyword are returned, has also been studied. However, in the setting of privacy assured similarity research, only single keyword is allowed, restricting the practical use.

III.PROBLEM ANALYSIS

Recently, privacy-preserving keyword search, such as secure ranked search, where only the files with better matching to the input keywords are returned, has been studied in the settings of single keyword and multi-keyword. Nonetheless, in the setting of secure ranked search, only the number of keyword matches is concerned, and the similarity between the input keywords and the actual words in text is not taken into account. On a different front, privacy assured similarity search, where the files containing exactly the same keyword or containing similar keyword are returned, has also been studied. However, in the setting of privacy assured similarity research, only single keyword is allowed, restricting the practical use. Recently, privacy-preserving keyword search, such as secure ranked search, where only the files with better matching to the input keywords are returned, has been studied in the settings of single keyword and multi-keyword. Nonetheless, in the setting of secure ranked search, only the number of keyword matches is concerned, and the similarity between the input keywords and the actual words in text is not taken into account. On a different front, privacy assured similarity search, where the files containing exactly the same keyword or containing similar keyword are returned, has also been studied. However, in the setting of privacy assured similarity research, only single keyword is allowed, restricting the practical use.

In this paper, we focus on privacy-preserving multi-keyword similarity search (PPMKSS) over the outsourced cloud data.

In the PPMKSS over the outsourced cloud data, the data are encrypted and then outsourced to the cloud. The user, after gaining the authorization, sends keywords to the cloud, which returns to the user the files containing as many keywords or their variants as possibleAll of the files containing at least a threshold number of keywords similar to the input keyword specified by the user will be returned to the user.

The cloud cannot learn additional information from the outsourced encrypted data and the corresponding index.

PPMKSS over the outsourced cloud data is considered for the first time in the literature.

The proposed system is a user authorization scheme with the guarantee of user access privacy by using blind signature.

By taking advantage of keyword suppressing technique and the Bloom filter, we propose three PPMKSS solutions, namely, PPMKSS-1, PPMKSS-2, and PPMKSS-3, to achieve PPMKSS.

These modules serves as a basis for the implementation of the project. Every module in the project lays steps for the good execution of the project. Every module specify the way to how to provide control over data sharing across the cloud. Description of modules are as follows.

- User Authorization
- PPMKSS-1
- PPMKSS-2
- PPMKSS-3

USER AUTHORIZATION

In this module, the user may access the files in the cloud by presenting the authorization token to the cloud first. More precisely, for the access control policy, we assume that once the user presents an authorization token to the cloud, the user is allowed to perform PPMKSS once. Moreover, after a particular authorization token is used for

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

a search request, it is no longer valid. To perform another keyword search, the user is required to present another authorization token.

To achieve the user access privacy, we propose to use blind signature protocol, which itself depends on the well-known Rivest–Shamir–Adleman (RSA) cryptosystem. The use of blind signature can have the guarantee that the owner signs and the cloud verify the authorization token without knowing the user's identity.

PPMKSS-1:

Basic Idea: Inverted list is a fundamental technique to support keyword search. We thus still follow the notion of inverted list to have our design for similarity search. Inverted list can be considered a 2-D array, each row of which consists of a tuple. The length of the array is equivalent to the number keywords that can be searched in the system. The first entry of each tuple is a keyword, whereas the second entry of each tuple consists of the file IDs that the corresponding files contain the keyword.

Protocol Description: The proposed PPMKSS-1 method is there are two phases: preprocessing and searching. During the preprocessing phase, the owner constructs and sends out the index for the further searching purpose. During the searching phase, the authorized user sends the keywords to the cloud, and then the cloud returns the search result to the user. The aforementioned method of constructing a very large inverted list is theoretically useful; however, it is practically useless. This is because the storage cost is overwhelming. For example, consider the keyword CLOUD.

PPMKSS-2:

Basic Idea: Instead of using the inverted list, in PPMKSS-2, the owner uses a more direct way to do the keyword search. In particular, once the keywords are selected and associated with the file, when the input keywords are sent to the cloud, the cloud is able to find out the files containing the input keywords. However, this straightforward method implies two problems. First, since there would be a large number of keyword variants, directly comparing the keywords would be a heavy computation cost for the cloud. Thus, we propose to use the Bloom filter to store the keyword variant and at the same time offer a more efficient membership query at the cloud side. However, the induced problem from the use of the Bloom filter is that, because the Bloom filter represents the keyword variants, the cloud may also query the Bloom filter to know the keywords associated with the particular files, breaching the keyword privacy. To solve this problem, we propose to use different sets of hash functions to disable the cloud's ability to check the keyword existence in files.

Protocol Description: The proposed PPMKSS-2 method is shown in Fig. 8. Before outsourcing the files to the cloud, the owner constructs a table, in which the first column consists of the file IDs, whereas the second column consists of the keywords and the keyword variants extracted from the corresponding. As can easily be known, such setting is completely opposite to the case in Section IV-B. Here, keyword variants are associated with the files, whereas in PPMKSS-1, the file IDs are associated with the keywords. As we can know from information retrieval literature, the setting in this section is more inefficient during the file retrieval.

PPMKSS-3:

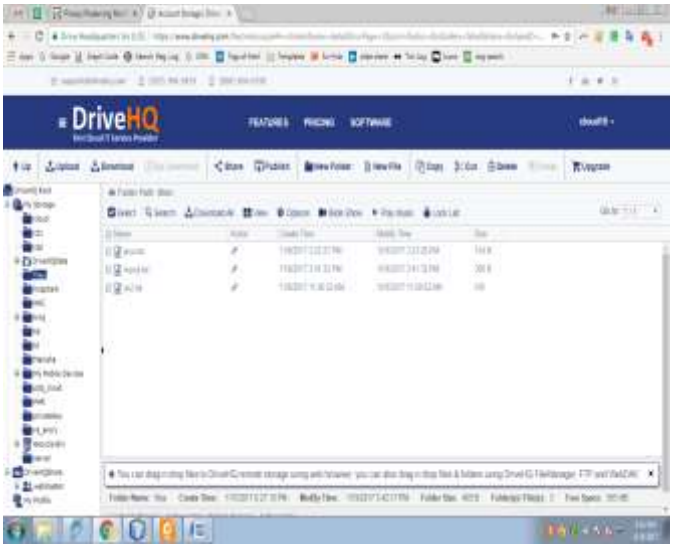
The PPMKSS-2 method, although theoretically feasible, is very inefficient in terms of search time because the cloud needs to access each file to make sure whether the particular files contain the input keyword variants. Thus, in this section, we combine the idea about the encrypted inverted list in Section IV-B and the idea about the Bloom table in Section IV-C to construct a new data structure to support the PPMKSS.

In this sense, the first column of the Bloom inverted list is completely the same as the first column of the encrypted inverted list in Section IV-B. Nevertheless, the file IDs in the inverted list are all added to a Bloom filter. As a consequence, for a Bloom inverted list, each keyword variant is associated with a Bloom filter.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

However, if the Bloom inverted list is directly outsourced to the cloud, the cloud is also able to test the keyword existence for each file by simply launching the membership query. By controlling the number of dummy elements, the Bloom inverted list is secure against the cloud attempting to know the keyword existence in files. Finally, at the end of the preprocessing phase, the owner sends the Bloom inverted list to the cloud.

IV.RESULT AND ANALYSIS



Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 3, March : 2023

CONCLUSION

Suppose consider the problem of PPMKSS over outsourced cloud data, for the first time in the literature. With the keyword suppressing technique and the Bloom filter, three solutions, namely, PPMKSS-1, PPMKSS-2, and PPMKSS-3, are proposed as candidates for dealing with such search problem. In particular, PPMKSS-3 is highly efficient in terms of storage, computation, and communication overhead. Moreover, we also design a user authorization mechanism based on blind signature, to ensure the user access privacy. As a whole, based on our evaluation, the proposed schemes can be practically useful in offering PPMKSS.

REFERENCES

[1] F.-H. Liu, H.-F. Lo, L.-C. Chen, and W.-T. Lee, "Comprehensive security integrated model and ontology within cloud computing," J. Internet Technol., vol. 14, no. 6, pp. 935–946, 2013.
[2] M. J. Atallah, Algorithms and Theory of Computation Handbook. Boca Raton, FL, USA: CRC Press, 1998.
[3] M. Armbrustet al., "A view of cloud computing," Commun. ACM, vol. 53, no. 4, pp. 50–58, Apr. 2010.
[4] B. Bloom, "Space/time trade-offs in hash coding with allowable errors," Commun. ACM, vol. 13, no. 7, pp. 422–426, Jul. 1970.
[5] B. Schneier, Applied Cryptography: Protocols, Algorithms, Source Code in C. Hoboken, NJ, USA: Wiley, 1996.
[6] M. Bellare, A. Boldyreva, and A. O'Neill, "Deterministic and efficiently searchable encryption," in Proc. Annu. CRYPTO, 2007, pp. 535–552.
[7] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. Int. Conf. Theory Appl. EUROCRYPT, 2004, pp. 506–522.
[8] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Proc. ICICS, 2005, pp. 414–426.
[9] A. Broder and M. Mitzenmacher, "Network applications of bloom filters: A survey," Internet Math., vol. 1, pp. 485–509, 2005.
[10] D. Boneh and B.Waters, "Conjunctive, subset, range queries on encrypted data," in Proc. Theory of Cryptography Conf. TCC, 2007, pp. 535–554.
[11] D. Chaum, "Security without identification: Transaction systems to make big brother obsolete," Commun. ACM, vol. 28, no. 10, pp. 1030–1044, Oct. 1985.

[12] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proc. ACM Conf. CCS, 2006, pp. 79–88.
[13] D. Cash et al., "Dynamic searchable encryption in very-Large databases: Data structures and implementation," in Proc. NDSS, 2014, pp. 1–16.
[14] Y.-C. Chang andM.Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. ACNS, 2005, pp. 442–455.
[15] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM, 2011, pp. 829–837.
[16] E.-J. Goh, "Secure indexes," Cryptology ePrint Archive, Santa Barbara, CA, USA, Rep. No. 2003/216, 2004.
[17] P. Golle, J. Staddon, and B. R. Waters, "Secure conjunctive keyword search over encrypted data," in Proc. ACNS, 2004, pp. 31–45.
[18] M. Kargar, A. An, N. Cercone, P. Godfrey, and J. Szlichta, "MeanKS: Meaningful keyword search in relational databases with complex schema," in Proc. ACM Int. Conf. SIGMOD, 2014, pp. 905–908.
[19] S. Kamara and K. Lauter, "Cryptographic cloud storage," in Proc. Financial Cryptography Workshops, Jan. 2010, pp. 136–149.