Catch You If You Misbehave: Ranked Keyword Search Results Verification In Cloud Computing

Kolanupaka. Dijili Saritha, Sri.S.K.Alisha, Sri.V.Bhaskara Murthy
Mca Student, Senior Assistant Professor, Associate Professor
Dept Of Mca
B.V.Raju College, Bhimavaram

ABSTRACT

With the advent of cloud computing, more and more people tend to outsource their data to the cloud. As fundamental data utilization, secure keyword search over encrypted cloud data has attracted the interest of many researchers recently. However, most of existing researches are based on an ideal assumption that the cloud server is "curious but honest", where the search results are not verified. In this paper, we consider a more challenging model, where the cloud server would probably behave dishonestly. Based on this model, we explore the problem of result verification for the secure ranked keyword search. Different from previous data verification schemes, we propose a novel deterrent-based scheme. With our carefully devised verification data, the cloud server cannot know which data owners, or how many data owners exchange anchor data which will be used for verifying the cloud server's misbehavior. With our systematically designed verification construction, the cloud server cannot know which data owners' data are embedded in the verification data buffer, or how many data owners' verification data are actually used for verification. All the cloud server knows is that, once he behaves dishonestly, he would be discovered with a high probability, and punished seriously once discovered. Furthermore, we propose to optimize the value of parameters used in the construction of the secret verification data buffer. Finally, with thorough analysis and extensive experiments, we confirm the efficacy and efficiency of our proposed schemes.

I. INTRODUCTION

With the advent of cloud computing, more and more people tend to outsource their data to the cloud. Cloud computing provides tremendous benefits including easy access, decreased costs, quick deployment, and flexible resource management [1], [2]. Enterprises of all sizes can leverage the cloud to increase innovation and collaboration.

Although cloud computing brings a lot of benefits, for privacy concerns, individuals and enterprise users are reluctant to outsource their sensitive data, including private photos, personal health records, and commercial confidential documents, to the cloud. Because once sensitive data are outsourced to a remote cloud, the corresponding data owner directly loses control of these data. The Apple's iCloud leakage of celebrity photo in 2014 [3] has furthered our concern regarding the cloud's data security. Encryption on sensitive data before outsourcing is an alternative way to preserve data privacy against adversaries. However, data encryption becomes an obstacle to the utilization of traditional applications,e.g., plaintext based keyword search.

To achieve efficient data retrieval from encrypted data, many researchers have recently

put efforts on secure keyword search over encrypted data [4], [5], [6], [7],[8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. However, all these schemes are based on the ideal assumption that the cloud server is "curious but honest". Unfortunately, in practical applications, the cloud server may be compromised and behave dishonestly [19], [20]. A compromised cloud server would return false search results to data users for various reasons:

1) The cloud server may return forged search results. For example, the cloud may rank an advertisement higher than others, since the cloud can profit from it, or the cloud would return random large files to earn money, since the cloud adopts the 'pay as you consume' model.

2) The cloud server may return incomplete search results in peak hours to avoid suffering from performance bottlenecks. There are some researches that focus on search results verification [21], [22], [23], [24], [25], [26], [27], [28], [29]. However, these methods cannot be applied to verify the top-k ranked search results in the cloud computing environment, where numerous data owners are involved.

We illustrate the reason from two aspects.

1) Existing schemes share a common assumption, i.e., data owners foresee the order of search results. However, in practical applications, numerous data owners are involved; each data owner only knows its own partial order.Without knowing the total order, these data owners cannot use the conventional schemes to verify the search results.

2) For a top-k ranked keyword search (e.g., k = 10), only a few data owners will have satisfied files in the search results. Traditional methods need to return a lot of data to verify whether the huge amount of absent data owners have satisfied search results. However, the top-k ranked keyword search is, to some extent, proposed to save communication cost; returning too much verification data would make the top-k ranked search meaningless. Additionally, in the 'pay as you consume' cloud computing environment, returning too much data would cause considerable expenses for data users, which would make the cloud computing lose its attractiveness.

In this paper, we consider a more challenging model, where multiple data owners are involved, and the cloud server would probably behave dishonestly. Based on this model, we explore the problem of result verification for the secure ranked keyword search. Different from previous data verification schemes, we propose a novel deterrent-based scheme. With our carefully devised verification data, the cloud server cannot know which data owners, or how many data owners exchange anchor data which will be used for verifying the cloud server's misbehavior. With our systematically designed verification construction, the cloud server cannot know which data owners' data are embedded in the verification data buffer, or how many data owners' verification data are actually used for verification. All the cloud server knows is that, once he behaves dishonestly, he would be discovered with a high probability, and punished seriously once discovered. Additionally, when any suspicious action is detected, data owners can dynamically update the verification data stored on the cloud server. Furthermore, we propose to optimize the value of parameters used in the construction of the secret verification data buffer. Finally, with thorough analysis and extensive experiments,

we confirm the efficacy and efficiency of our proposed schemes.

The main contributions of this paper are:

We formalize the ranked keyword search result verification problem where multiple data owners

are involved and the cloud server would probably behave dishonestly.

We propose a novel secure and efficient deterrentbased verification scheme for secure ranked keyword search.

We propose to optimize the value of parameters used in the construction of verification data buffer.

We give a thorough analysis and conduct extensive performance experiments to show the efficacy and efficiency of our proposed scheme.

The rest of this paper is organized as follows. Section 2 reviews the related works. Section 3 formulates the problem and introduces notations used in later discussions. Section 4 describes the preliminary techniques that will be used in this paper. In Section 5, we illustrate how to efficiently and securely verify the ranked search results. In Section 6, we show how to optimize the value of parameters. In Sections 7 and 8, we present analysis and performance evaluation for our proposed schemes, respectively.

II. EXISTING SYSTEM

However, most of existing researches are based on an ideal assumption that the cloud server is “curious but honest”, where the search results are not verified. In this paper, we consider a more challenging model, where the cloud server would probably behave dishonestly. Based on this model, we explore the problem of result verification for the secure ranked keyword search. Different from previous data verification schemes, we propose a novel deterrent-based scheme. With our carefully devised verification data, the cloud server cannot know which data owners, or how many data owners exchange anchor data which will be used for verifying the cloud server’s misbehavior.

III. PROPOSED SYSTEM

Furthermore, we propose to optimize the value of parameters used in the construction of the secret verification data buffer. Finally, with thorough analysis and extensive experiments, we confirm the efficacy and efficiency of our proposed schemes.

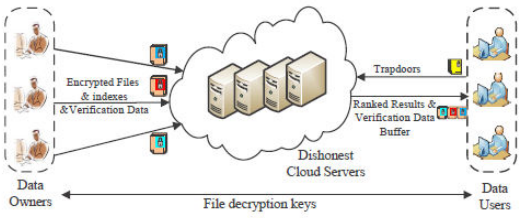
1. Proposed to save communication cost;

Returning too much verification data would make the top-k ranked search meaningless. Additionally, in the ‘pay as you consume’ cloud computing environment, returning too much data would cause considerable expenses for data users, which would make the cloud computing lose its attractiveness.

2. The main contributions of this paper are:

We formalize the ranked keyword search result verification problem where multiple data owners are involved and the cloud server would probably behave dishonestly. We propose a novel secure and efficient deterrentbased verification scheme for secure ranked keyword search. We propose to optimize the value of parameters used in the construction of verification data buffer. We give a thorough analysis and conduct extensive performance experiments to show the efficacy and efficiency of our proposed scheme.

IV. SYSTEM ARCHITECTURE



V. IMPLEMENTATION MODULES

- Secure Keyword Search in Cloud Computing.
- Verifying Ranked Top-k Search Results
- Privacy Preserving Ranked Keyword Search Among Multiple Data Owners.
- Assembling the verification data.
- Returning verification data

MODULE DESCRIPTION

Secure Keyword Search in Cloud Computing:

Recently, there have been a lot of research works concerned with secure keyword search in cloud computing. The first securely ranked keyword search over encrypted data was proposed by Wang et al.. Cao et al. and Wen et al. further strengthening the ranked keyword search and constructing schemes for privacypreserving multi-keyword ranked search. In, Xu et al. proposed a multi-keyword ranked query scheme on encrypted data, which enables a dynamic keyword dictionary and avoids the problem in which the rank order is perturbed by several high frequency keywords. Based on information retrieval systems and cryptography approaches, Ibrahim et al. proposed a ranked searchable encryption scheme of multi-keyword search over a cloud server. Hore et al. further proposed using a set of colors to encode the presence of the keywords and creating an index to accelerate the search process.

Verifying Ranked Top-K Search Results

The basic idea of our deterrent based verification scheme is elaborated as follows: We can consider the dishonest cloud server as a suspect, the data user as a police chief, and each verification data as a policeman, who masters part of the suspect's actions. Intuitively, the police chief can gather all the policemen to verify whether the suspect commits a crime. However, this will cause a lot of manpower, financial and time waste. To overcome this problem, each time the suspect takes an action, the police chief only inquires a few policemen to verify whether the suspect commits a crime. During the process, the police chief ensures that the suspect does not know which policemen know his action, and which policemen are inquired by the police chief. What the suspect knows is that, once he behaves dishonestly, he will be discovered with high probability, and punished seriously once discovered. By doing this, we can deter the suspect not to behave dishonestly.

Privacy Preserving Ranked Keyword Search Among Multiple Data Owners:

In our previous work , we introduce how to achieve ranked and privacy-preserving keyword search among multiple data owners. First of all, we systematically construct protocols on how to encrypt keywords for data owners, how to generate trapdoors for data users, and how to perform blind searching for the cloud server. As a result, different data owners use their own secret keys to encrypt their files and keywords. Authorized data users can issue queries without knowing secret keys of these data owners. Then an Additive Order Preserving Function family is proposed, which enables different data owners to encode their relevance scores with different secret keys, and helps cloud server return the top-k relevant search results to data users without revealing any sensitive information. In this paper, we adopt this ranked and privacy preserving keyword search scheme to return the top-k search results. Our goal is to systematically construct schemes that can verify whether the returned top-k search results are correct.

Assembling the verification data:

When an authorized data user wants to verify the search results, he specifies a set of data owners whose verification data need to be returned to help verification. The data user can achieve this goal by simply setting an ID set of his desired data owners. However, the ID set should not be exposed to the cloud server. The fundamental reason is illustrated as follows: if the cloud server knows which data owners' data are frequently verified, he can deduce that these data owners' data are very useful or sensitive, therefore, these data owners' data would easily become attackers' targets. On the other hand, if the cloud server knows which data owners' data are rarely verified, the cloud server will maliciously filter out or delete these data owners' data as search results. To prevent the cloud server from knowing which data owners' data are actually returned, we propose to construct a secret verification request which is illustrated as follows: First, the data user enlarges the ID set of verification by inserting random IDs. Assume a data user wants to get Oi's verification data, he can add other n−1 data owners' ID in the set (we can adopt encryption or obfuscation to hide the true ID, for easy description, we simply demonstrate with ID hereafter). Second, the data user attaches a data 0 or 1 to each ID. Here, if the data user wants to return a data owner's verification data, then he attaches 1 to the corresponding ID, otherwise, 0 is attached. Third, the data user encrypts the attached 0 or 1 with the Paillier encryption

Returning verification data:

When the data user gets some data owners' verification data, he can further recover all the sampled data and anchor data. The data user will use them to verify whether the returned results are correct. The verification is done in two steps: first, the data user verifies whether the data from a specific data owner is correct. If the search results pass the first verification, the verification process turns to the second step, i.e., with the help of anchor data, the data user verifies whether the search results from different data owners are correct. After verification, the data user can detect the cloud server's misbehavior with a high probability. In Section , we will give an analysis of the detection probability

VI. CONCLUSION

In this paper, we explore the problem of verification for the secure ranked keyword search, under the model where cloud servers would probably behave dishonestly. Different from previous data verification schemes, we propose a novel deterrent-based scheme. During the whole process of verification, the cloud server is not clear of which data owners, or how many data owners exchange anchor data used for verification, he also does not know which data owners' data are embedded in the verification data buffer or how many data owners' verification data are actually used for verification. All the cloud server knows is that, once he behaves dishonestly, he would be discovered with a high probability, and punished seriously once discovered. Additionally, when any suspicious action is detected, data owners can dynamically update the verification data stored on the cloud server. Furthermore, our proposed scheme allows the data users to control the communication cost for the verification according to their preferences, which is especially important for the resource limited data users. Finally, with thorough analysis and extensive experiments, we confirm the efficacy and efficiency of our proposed schemes.

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. Zhu, V. Leung, X. Hu, L. Shu, and L. T. Yang, "A review of key issues that concern the feasibility of mobile cloud computing," in Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing. IEEE, 2013, pp.769–776.

[3] Ritz, "Vulnerable icloud may be the reason to celebrity photo leak." [Online]. Available: http://marcritz.com/icloud-flaw-leak/

[4] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE Distributed Computing Systems (ICDCS'10), Genoa, Italy, Jun. 2010, pp. 253–262.

[5] 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'11, Shanghai, China, Apr. 2011, pp. 829–837.

[6] W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking," in Proc. IEEE ASIACCS'13, Hangzhou, China, May 2013, pp. 71–81.

[7] Z. Xu,W. Kang, R. Li, K. Yow, and C. Xu, "Efficient multi-keyword ranked query on encrypted data in the cloud," in Proc. IEEE Parallel and Distributed Systems (ICPADS'12), Singapore, Dec. 2012,pp. 244–251.

[8] A. Ibrahim, H. Jin, A. A. Yassin, and D. Zou, "Secure rankordered search of multi-keyword trapdoor over encrypted cloud data," in Proc. IEEE Asia-Pacific Conference on Services Computing(APSCC'12), Guilin, China, Dec. 2012, pp. 263–270.

[9] B. Hore, E. C. Chang, M. H. Diallo, and S. Mehrotra, "Indexing encrypted documents for supporting efficient keyword search," in Proc. Secure Data Management (SDM'12), Istanbul, Turkey, Aug. 2012, pp. 93–110.

[10] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. IEEE INFOCOM'10, San Diego, CA, Mar. 2010, pp. 1–5.

[11] M. Chuah and W. Hu, "Privacy-aware bedtree based solution for fuzzy multi-keyword search over encrypted data," in Proc. IEEE 31th International Conference on Distributed Computing Systems(ICDCS'11), Minneapolis, MN, Jun. 2011, pp. 383–392.

[12] P. Xu, H. Jin, Q. Wu, and W. Wang, "Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack," IEEE Transactions on Computers, vol. 62, no. 11,pp. 2266–2277, 2013.

[13] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multikeyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, Toronto, Canada, May 2014, pp. 2112–2120.

[14] C. Wang, K. Ren, S. Yu, and K. M. R. Urs, "Achieving usable and privacy-assured similarity search over outsourced cloud data," in Proc. IEEE INFOCOM'12, Orlando, FL, Mar. 2012, pp. 451–459.

[15] W. Sun, S. Yu, W. Lou, Y. T. Hou, and H. Li, "Protecting your right: Attribute-based keyword search with fine-grained ownerenforced search authorization in the cloud," in Proc. IEEE INFOCOM' 14, Toronto, Canada, May 2014, pp. 226–234.

[16] Q. Zheng, S. Xu, and G. Ateniese, “Vabks: Verifiable attributebased keyword search over outsourced encrypted data,” in Proc. IEEE INFOCOM’14, Toronto, Canada, May 2014, pp. 522–530.

[17] W. Zhang, S. Xiao, Y. Lin, T. Zhou, and S. Zhou, “Secure ranked multi-keyword search for multiple data owners in cloud computing,” in Proc. 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN2014). Atlanta, USA: IEEE, jun 2014, pp. 276–286.

[18] W. Zhang, Y. Lin, S. Xiao, Q. Liu, and T. Zhou, “Secure distributed keyword search in multiple clouds,” in Proc. IEEE/ACM IWQOS’14, Hongkong, May 2014, pp. 370–379.

[19] B. Wang, B. Li, and H. Li, “Oruta: Privacy-preserving public auditing for shared data in the cloud,” IEEE Transactions on Cloud Computing, vol. 2, no. 1, pp. 43–56, 2014.

[20] J. Li, X. Tan, X. Chen, D. Wong, and F. Xhafa, “Opor: En-abling proof of retrievability in cloud computing with resourceconstrained devices,” IEEE Transactions on Cloud Computing, 2014.