SECURITY FOR ENCRYPTED CLOUD DATA BY USING TOP-KEY TREE TECHNOLOGIES

[1]MANJOORULLASHA SHAIK, [2]SYED.ABDULHAQ,[3] P.BABU

[1]PG SCHOLAR, CSE (CN), QCET, NELLORE
[2,3]ASSOCIATE PROFESSOR, CSE, QCET, NELLORE

ABSTRACT: Cloud computing has emerging as a promising pattern for data outsourcing and high quality data services. However, concerns of sensitive information on cloud potentially cause privacy problems. Data encryption protects data security to some extent, but at the cost of compromised efficiency. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud. In this paper, we focus on addressing data privacy issues using searchable symmetric encryption (SSE). For the first time, we formulate the privacy isue from the aspect of similarity relevance and scheme robustness. We observe that server-side ranking based on order-preserving encryption (OPE) inevitably leaks data privacy. To eliminate the leakage, we propose a two-round searchable encryption (TRSE) scheme that supports top-k multi-keyword retrieval. In TRSE, we employ a vector space model and homomorphic encryption. The vector space model helps to provide sufficient search accuracy, and the homomorphic encryption enables users to involve in the ranking while the majority of computing work is done on the server side by operations only on ciphertext. As a result, information leakage can be eliminated and data security is ensured. Thorough security and performance analysis show that the proposed scheme guarantees high security and practical efficiency.

1. INTRODUCTION: Cloud computing a critical pattern for advanced data service, has became a necessary feasibility for data users to outsource data. Controversies on privacy, however, have been incessantly presented as outsourcing of sensitive information including emails, health history and personal photos is explosively expanding. Reports of data loss and privacy breaches in cloud computing systems appear from time to time . The main threat on data privacy roots in the cloud itself 6. When users outsource their private data onto the cloud, the cloud service providers are able to control and monitor the data and the communication between users and the cloud at will, lawfully or unlawfully,. Instances such as the secret NSA program, working with AT&T and Verizon, which recorded over million phone calls between American citizens, cause uncertainty among privacy advocates, and the greater powers it gives to telecommunication companies to monitor user activity 7. To ensure privacy, users usually encrypt the data before outsourcing it onto cloud, which brings great challenges to effective data utilization. However, even if the encrypted data utilization is possible, users still need to communicate with the cloud and allow the cloud operates on the encrypted data, which potentially causes leakage of sensitive information. Furthermore, in cloud computing, data owners may share their outsourced data with a number of users, who might want to only retrieve the data files they are interested in. One of the most popular ways to do so is through keyword-based retrieval. Keyword-based retrieval is a typical data service and widely applied in plaintext scenarios, in which users retrieve relevant files in a file set based on keywords. However, it turns out to be a difficult task in ciphertext scenario due to limited operations on encrypted data. Besides, in order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevances are sent back

to users. A series of searchable symmetric encryption schemes have been proposed to enable search on ciphertext. Traditional SSE schemes enable users to securely retrieve the ciphertext, but these schemes support only boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. To improve security without sacrificing efficiency, schemes presented in 91024 show that they support top-k single keyword retrieval under various scenarios. Authors of made attempts to solve the problem of top-k multi-keyword over encrypted cloud data. These schemes, however, suffer from two problems - boolean representation and how to strike a balance between security and efficiency. In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy. In the latter, security is implicitly compromised to tradeoff for efficiency, which is particularly undesirable in security-oriented applications. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage. However, the limited computational power on the user side and the high computational overhead precludes information security. The issue of secure multi-keyword top-k retrieval over encrypted cloud data thus is: how to make the cloud do more work during the process of retrieval without information leakage. In this paper, we introduce the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval community are employed, including homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees topk multi-keyword retrieval over encrypted cloud data with high security and practical efficiency. Our contributions can be summarized as follows:

2. EXISTING SYSTEM

Besides, in order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevances are sent back to users. A series of searchable symmetric encryption schemes have been proposed to enable search on ciphertext. Traditional SSE schemes enable users to securely retrieve the ciphertext, but these schemes support only Boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage. However, the limited computational power on the user side and the high computational overhead precludes information security.

2.1 DISADVANTAGES OF EXISTING SYSTEM:

- To improve security without sacrificing efficiency, schemes presented in show that they support top-k single keyword retrieval under various scenarios.
- Authors of made attempts to solve the problem of top-k multi-keyword over encrypted cloud data.
- These schemes, however, suffer from two problems - Boolean representation and how to strike a balance between security and efficiency.

- ✖ In the former, files are ranked only by the number of retrieved keywords, which impairs search accuracy. In the latter, security is implicitly compromised to tradeoff for efficiency, which is particularly undesirable in security-oriented applications.

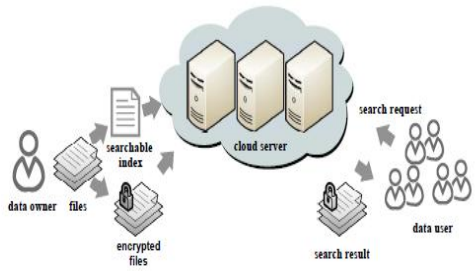
3. PROPOSED SYSTEM:

In this paper, we introduce the concepts of similarity relevance and scheme robustness to formulate the privacy issue in searchable encryption schemes, and then solve the insecurity problem by proposing a two-round searchable encryption (TRSE) scheme. Novel technologies in the cryptography community and information retrieval community are employed, including homomorphic encryption and vector space model. In the proposed scheme, the majority of computing work is done on the cloud while the user takes part in ranking, which guarantees top k multi-keyword retrieval over encrypted cloud data with high security and practical efficiency.

3.1 ADVANTAGES OF PROPOSED SYSTEM:

- ✓ We propose the concepts of similarity relevance and scheme robustness. We thus perform the first attempt to formulate the privacy issue in searchable encryption, and we show server side ranking based on order-preserving encryption (OPE) inevitably violates data privacy
- ✓ We propose a two-round searchable encryption (TRSE) scheme, which fulfills the secure multi-keyword top-k retrieval over encrypted cloud data. Specifically, for the first time we employ relevance score to support multi-keyword top-k retrieval.
- ✓ Thorough analysis on security demonstrates the proposed scheme guarantees high data privacy. Furthermore, performance analysis and experimental results show that our scheme is efficient for practical utilization.

4. SYSTEM ARCHITECTURE:



4.1 ALGORITHM USED:

Algorithm 1 TOPKSELECT(*source*, *k*)

Input:

list *source* to be selected

number *k*

Initialization:

Set $topk \leftarrow \emptyset; topkid \leftarrow \emptyset;$

Iteration:

1: **for all** $item \in source$ **do**
2: INSERT(*topk*, (*item*, *itemindex*))
3: **end for**
4: **for all** $tuple \in topk$ **do**
5: *topkid.append*(*tuple*[1])
6: **end for**

Output:

topkid

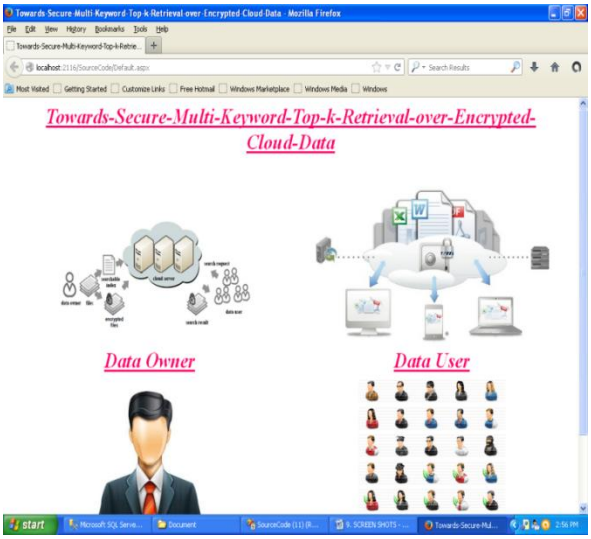
Algorithm 2 INSERT$(topk, (item, itemindex))$

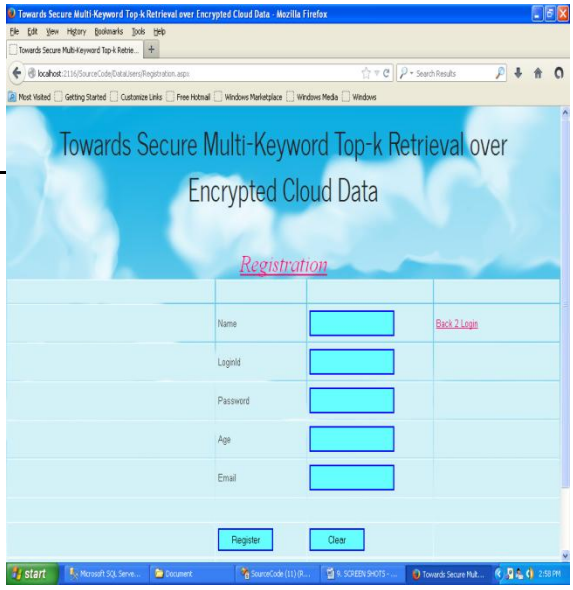
Input:

list $topk$ to store the top-k scoring item tuple $(item, itemindex)$

Iteration:

```
1: if len(topk) < k then
2:     insert (item, itemindex) into topk in nondecr
       order of item
3: else
4:     for all element ∈ topk do
5:         if item < element[0] then
6:             continue
7:         else
8:             discard topk[0], insert (item, itemindex)
               topk in nondecreasing order of item
9:         end if
10:    end for
11: end if
```





SYSTEM REQUIREMENTS:

HARDWARE REQUIREMENTS:

- System : Pentium IV 2.4 GHz.
- Hard Disk : 40 GB.
- Monitor : 15inch VGA Colour.
- Mouse : Logitech Mouse.
- Ram : 512 MB
- Keyboard : Standard Keyboard

SOFTWARE REQUIREMENTS:

- Operating System:Windows XP.
- Coding Language:ASP.NET, C#.Net.
- Database : SQL Server 2005

5.CONCLUSION:

In this paper, we motivate and solve the problem of secure multi-keyword top-k retrieval over encrypted cloud data. we define similarity relevance and scheme robustness. Based on orderpreserving encryption invisibly leak sensitive information, we devise a server-side ranking SSE scheme. We then propose a two-round searchable encryption (TRSE) scheme employing the fully homomorphic encryption, which fulfills the security requirements of multi-keyword topk retrieval over the encrypted cloud data. By security analysis, we show that the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over real dataset, extensive experimental results

demonstrate that our scheme ensures practical efficiency.

REFERENCES:

[1] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia. "A view of cloud computing," Communication of the ACM 53 (4): 50 58, 2010.

[2] M. Arrington, "Gmail disaster: Reports of mass email deletions," http://www.techcrunch.com/2006/12/28/gmail-disasterreports-of-mass-email-deletions/, December 2006.

[3] Amazon.com, "Amazon s3 availability event: July 20, 2008," http://status.aws.amazon.com/s3-20080720.html, 2008.

[4] RAWA News, "Massive information leak shakes Washington over Afghan war," http://www.rawa.org/temp/runews/2010/08/20/massive-information-leak-shakeswashington- over-afghan-war.html, 2010

[5] AHN, "Romney hits Obama for security information leakage," http://gantdaily.com/2012/07/25/romney-hits-obama-for-security-information-leakage/, 2012

[6] Cloud Security Alliance, "Top threats to cloud computing," http://www.cloudsecurity alliance. org, 2010.

[7] C. Leslie, "NSA has massive database of Americans' phone calls," http://usatoday30.usatoday.com/news/washington/2006-05-10/.

[8] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," in Proc. of ACM CCS, 2006.

[9] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS, 2010.

[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+r: Top-k retrieval from a confidential index," in Proc. of EDBT, 2009.

[11] M. van Dijk, C. Gentry, S. Halevi and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers," in Gilbert, H. (ed.) EUROCRYPT. LNCS, vol. 6110, pp. 24-43, 2010.

[12] M. Perc, "Evolution of the most common English words and phrases over the centuries," the Journal of the Royal Society Interface, 2012. / mec/2003-2004/.

[13] O. Regev, "New lattice-based cryptographic constructions," JACM 51(6), pp. 899-942, 2004.

[14] N. Howgrave-Graham, "Approximate integer common divisors," in Silverman, J.H. (ed.) CaLC' 01. LNCS, vol. 2146, pp. 51-66, 2001.

[15] NSF Research Awards Abstracts 1990-2003: http://kdd.ics.uci.edu/databases/nsfabs /nsfawards