ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

IJIRSET

International Journal of Innovative Research in Science, Engineering and Technology

An ISO 3297: 2007 Certified Organization Volume 3, Special Issue 5, July 2014

International Conference On Innovations & Advances In Science, Engineering And Technology [IC - IASET 2014]

Organized by

Toc H Institute of Science & Technology, Arakunnam, Kerala, India during 16th -18th July -2014

Enabling Search and Retrieval over Encrypted Data Using Homomorphic Encryption

Jiss Varghese[1], Lisha Varghese[2], Fabeela Ali Rawther [3]

PG Scholar, Dept of Computer Science, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India[1]

Assistant Professor, Dept of Computer Science, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India[2]

Assistant Professor, Dept of Computer Science, Amal Jyothi College of Engineering, Kanjirappally, Kerala, India[3]

ABSTRACT: Outsourcing of data has become an advanced data service to store information into a global storage space. Data owners can upload their private data onto these space, and do many operations on them. Performing information retrieval tasks while preserving data confidentiality is a big challenge when data is stored on a global space maintained by a third-party service provider. To avoid confidentiality problem, the sensitive data have to be encrypted before uploading onto servers. This leads to a problem to perform efficient keyword based search and retrieval over the encrypted data. Several searchable symmetric encryption schemes are available but they support only Boolean keyword based search. i.e., a file is retrieved based upon whether the queried keyword is present or not without considering the relevance. This paper proposes a novel scheme called homomorphic encryption which allows different computations to be applied on encrypted data. This property enables scoring and ranking computations which gives sufficient search accuracy and enables users to involve in ranking by doing computations only on ciphertext. Since search operation is performed over encrypted data, information leakage can be eliminated and data can be searched and retrieved efficiently.

KEYWORDS: Homomorphic encryption, searchable symmetric encryption, relevance, data retrieval

I. INTRODUCTION

As data outsourcing become more flexible and effective in data management, data owners are motivated to outsource their complex data from local space to global storage space. But for security of data, sensitive data has to be encrypted before outsourcing, which overcomes method of traditional data utilization based on plaintext keyword search. The goal of information retrieval over an encrypted data is to provide efficient and accurate search capability over encrypted documents without decrypting them first. Advancements in this area can have applications in protecting the privacy of sensitive data stored on third-party servers. Considering the large number of data users and documents, it is necessary for a search service to allow multi-keyword query and provide ranked result to meet the effective data retrieval need. Retrieving of all the files having queried keyword is not a good strategy. So some enhanced methods are required.

In order to protect data privacy, data need to be encrypted before being transferred to the remote server. Data encryption makes effective data utilization a very challenging task given that there could be a large amount of
outsourced data files. Besides, data owners may share their outsourced data with a large number of users, who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through keyword-based search. This keyword search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios. Unfortunately, data encryption, which restricts user's ability to perform plaintext keyword search and further demands the protection of keyword privacy, makes the traditional plaintext

search methods fail for encrypted data. Ranked search greatly improves system usability by normal matching files in a ranked order regarding to certain relevance criteria[6]. Encryption can be done using ciphers such as AES or RSA directly. Built upon the established cryptographic encryption tools, it is computationally difficult for the server to decrypt data files in order to learn the database content. Encryption keeps data content safe from the server but also makes it difficult for the server to build searchable indexes. In secure retrieval scenario, the search indexes need to be generated and properly encrypted by software tools on the user side using a secret key and then transferred to the server[7]. This paper proposes a novel scheme called homomorphic encryption which allows different computations to be applied on encrypted data[1]. This property enables scoring and ranking computations which gives sufficient search accuracy and enables users to involve in ranking by doing computations only on ciphertext. Since search operation is performed over encrypted data, information leakage can be eliminated and data can be searched and retrieved efficiently.

A basic architecture for this scenario is shown in Figure 1.

Fig. 1.1. Search and Retrieval Scenario

Consider a computing system hosting data management services. There are three different entities: Data owner, data user, and a server for storing the data files. Data owner has a collection of data files to be uploaded into server in encrypted format. The data owner, before sending data, will first build a homomorphically encrypted searchable index, and then outsource both the index and the encrypted files collection to server space. To search for a file, an authorized user should construct a search query, encrypt it homomorphically. Upon receiving it from data users, server is responsible to search the encrypted index, calculate the scores and returns the encrypted scores of corresponding set of encrypted files. Scoring is a natural way to weight the relevance of a document with respect to a query. Here tf-idf scoring scheme is used where term frequency and inverse document frequency determines the score value. Next on receiving the score values, data user decrypt the scores, rank them, picks out the top-k highest scoring files' identifiers and send them to server. Server then returns the top-k encrypted files. Users receive them, decrypt it and use the requested files. To improve file retrieval accuracy, search result should be ranked by user according to some ranking criteria. Server only sends back top-*k* files that are most relevant to the search query. If any Unauthorized user tries to access any data from the storage server then it is failed since only authorized users can make use of encrypted files.

The rest of this paper is organized as follows: Section II contains some related work, section III explains the proposed Homomorphic encryption mechanism, section IV deals with complexity analysis and, conclusion and future scope of work is given in section V.

II. RELATED WORK

Searchable symmetric encryption (SSE) allows a client to encrypt data in such a way that it can later perform search and retrieval from the storage server. Given a query, the server can search over the encrypted data and return the appropriate encrypted files. Informally, a SSE scheme is efficient if: (1) the ciphertext alone reveals no information about the data; (2) the ciphertext together with a search query reveals at most the result of the search; (3) search query can only be generated using the secret key.

Existing searchable encryption schemes[2],[4] allow a user to securely search over encrypted data through keywords without first decrypting it, these techniques support only conventional Boolean keyword search, without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing service, they go through following disadvantage. Most of the schemes follows either single keyword search or boolean keyword search without ranking and thereby do not get relevant data.

Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency). One simple ranked keyword search is implemented using the order-preserving symmetric encryption (OPSE). Order Preserving symmetric Encryption (OPE) [3] is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plaintexts. OPE not only allows efficient range queries, but allows indexing and query processing to be done exactly and as efficiently as for unencrypted data. The main drawback of OPE scheme is that it inevitably leaks data privacy. Eventhough data are in encrypted form the curious server or attacker can still obtain additional information through statistical analysis. The possible information leakage is termed as statistic leakage. There are two possible statistic leakages, including term distribution and inter distribution. The term distribution of term t is t's frequency distribution of scores on each file i (i € C). The inter distribution of file f is file f's frequency distribution of scores of each term j (j € f). For instance, obviously, the term “states” are very likely to co-occur with “united” in an official paperwork from the White House, and their term distribution, not surprisingly, are very same in a series of such a kind of paperwork. Given that this paperwork is encrypted but term distribution is not concealed, once an adversary somehow cracks out the plaintext of “united”, he can reasonably guess the term that shares a similar term distribution with “united” may be “states”.

Thus to improve security without sacrificing efficiency, a novel encryption mechanism called Homomorphic Encryption is proposed. Ranked search improves system usability by normal matching files in a ranked order regarding to certain relevance criteria (e.g., keyword frequency). As directly outsourcing relevance scores will drips a lot of sensitive information against the keyword privacy. This paper proposes a novel encryption with ranking result of queried data which will give only relevant data.

III. PROPOSED APPROACH

Here a fully homomorphic encryption scheme is proposed, solving a central open problem in cryptography. It can be defined as a form of encryption where a specific algebraic operation performed on the plaintext is equivalent to another (possibly different) algebraic operation performed on the cipher text[5]. An example shows how homomorphic encryption could be used to solve two numbers 5+3(which are in encrypted form). The data is encrypted locally, 5 becomes 55 and 3 becoming 33. The encrypted numbers are sent to the server where the encrypted numbers are added together. The server then returns the encrypted answer of 88. It is then decrypted locally to reveal the final answer of 8. The homomorphic encryption enables private queries to a search engine. The user submits an encrypted query and the search engine computes a succinct encrypted answer without ever looking at the query in the clear. It also enables searching on encrypted data. Consider the scenario. A user stores encrypted files on a remote file server and can later have the server retrieve only files that (when decrypted) satisfy some boolean constraint, even though the server cannot decrypt the files on its own. More broadly, fully homomorphic encryption improves the efficiency of secure multiparty computation.

Homomorphic encryption allows specific types of computations to be carried out on the corresponding ciphertext. The result is the ciphertext of the result of the same operations performed on the plaintext. i.e., homomorphic encryption allows computation of ciphertext without knowing anything about the plaintext to get the correct encrypted result. Only addition and multiplication operations over integers are needed to compute the relevance scores from the encrypted searchable index. Therefore, we can reduce the original homomorphism in a full form to a simplified form that only supports integer operations, which allows more simplicity than the full form does.

In the fully homomorphic encryption over the integers (FHEI) scheme [14][13], the approximate integer Greatest Common Divisor (GCD) is used to provide sufficient security, i.e., given a list of integers $l = \{i1, i2,\ldots in\}$ that are approximate multiples of a hidden integer j, to find the hidden integer j. The approximate GCD problem has been proven hard by Howgrave-Graham. Let m and c denote the plaintext and ciphertext of the integer, respectively. The encryption scheme can be expressed as the following formulation: Ciphertext $C = pq + 2r + m$ where m denotes the plaintext, p denotes the secret key, q denotes the multiple parameter, and r denotes the noise to achieve proximity against brute-force attacks. The public key is $pq + r$. On the basis of homomorphism property, the encryption scheme can be described as four stages[11]:

A. Key Generation

The secret key SK is an odd n-bit number randomly selected from the interval [2n-1, 2n). The secret key is used for encryption and the public keys are used for decryption, which are different from the concepts of keys in public-key cryptography.

B. Encryption

Ciphertext is given by $xr + m + ki$

C. Evaluation

Apply the binary addition and multiplication gates to the t ciphertext Ci, perform all necessary operations, and then return the resulting integer Ƙ.

D. Decryption

Output M = (Ƙ mod p) mod x

IV. ANALYSIS

As discussed before, to perform search and retrieval over encrypted data, several searchable symmetric encryption schemes are available. The Order preserving encryption is commonly used, but it causes information leakage. Eventhough data are in encrypted form the curious server or attacker can still obtain additional information through statistical analysis. The nature of access pattern and search pattern are revealed even the data files are in encrypted form. The new Homomorphic encryption do not have such problem. Since both addition and multiplication operations can be performed over encrypted data, scoring and ranking can be calculated over encrypted data. Thereby information leakage can be eliminated and data can be searched and retrieved efficiently. The attacker could not reveal the nature of plaintext since the ciphertext is a sequence of integers which do not have any statistical similarity with the plaintext.

A. Order Preserving Encryption

An OPE scheme is a deterministic symmetric-key encryption scheme that preserves the order of the plaintexts. It is not a perfectly secure encryption scheme since ciphertexts inevitably leak the order information of the plaintexts. for an OPE scheme as the expected number of bits of a plaintext remaining secret against the known plaintext attack. Security of OPE scheme can be derived by computing the average min-entropy of the plaintext given the ciphertext and known plaintext/ciphertext pairs. In entropy based analysis, security is defined as the number of information-theoretically secure bits of x against the known plaintext attack. Although the adversary may retrieve some information about the plaintext x, the probability for the adversary to fully recover the plaintext x is a negligible.

B. Homomorphic Encryption

According to the parameter selection in the Homomorphic Encryption scheme, the complexity is O(λ10). The parameter λ used here is a fixed integer. To build index I, several technologies from IR community can be used. The homomorphic encryption needs only the addition operation, so the complexity of encrypting I is O(nl), where n denotes the number of files and l denotes the number of keywords. The query-vector generation needs O(l) time to build the l-dimension query-vector from the multikeyword request. To encrypt this each dimension needs to be encrypted. Since the encryption requires only addition operations, the complexity of this stage is O(l).

V. CONCLUSION AND FUTURE SCOPE

The main problem in information retrieval scenario is that statistical information leakage is possible even the data are in encrypted form. The Order preserving encryption inevitably cause data leakage. Here a new scheme have been developed which enable data owner to upload encrypted data files into server and allow several authorised users to perform search and retrieval over these encrypted data. It enable user to get the retrieval result with the most relevant files that match users' interest instead of all the files. Data files are ranked in the order of relevance by users' interest and only the files with the highest relevance are sent back to users. Since statistical analysis is not possible in homomorphic ciphertext, the proposed scheme give security to certain extend. Users can do search and retrieval efficiently.

The developed scheme support single data owner and associated data users. The scheme can be extended by including provision for multiple data owners and associated set of data users. There will be separate space for each data owner at server side. Only authorised set of users can make use of data of corresponding data owners. Also key management should be included since maintaining keys is a major issue. Several researches in cryptographic community tries to improve the security and efficiency of homomorphic encryption[8]. This indicate that the efficiency of the TSSE scheme can be further improved.

REFERENCES

[1] Jiadi Yu, Member IEEE, Peng Lu, Yanmin Zhu, Member IEEE, Guangtao Xue, Member IEEE Computer Society, and Minglu Li "Toward Secure Multikeyword Top-k Retrieval over Encrypted Cloud Data", IEEE Transactions on Dependable and Secure Computing, Vol. 10, NO. 4, July/August 2013.
[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data", Proc. IEEE 30th Intl Conf. Distributed Computing Systems (ICDCS), 2010.
[3] Rakesh Agrawal, Jerry Kiernan, Ramakrishnan Srikant Yirong Xu , "Order Preserving Encryption for Numeric Data", IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120
[4] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions", Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.
[5] M. van Dijk, C. Gentry, S. Halevi, and V. Vaikuntanathan, "Fully Homomorphic Encryption over the Integers", Proc. 29th Ann. Intl Conf. Theory and Applications of Cryptographic Techniques, H. Gilbert, pp. 24-43, 2010.
[6] D. Song, D. Wagner, and A. Perrig, "Practical Techniques for Searches on Encrypted Data," Proc. IEEE Symp. Security and Privacy, 2000.
[7] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-Preserving Multikeyword Ranked Search over Encrypted Cloud Data," Proc. IEEE INFOCOM, 2011
[8] N. Smart and F. Vercauteren, "Fully Homomorphic Encryption with Relatively Small Key and Ciphertext Sizes," Proc. 13th Int'l Conf. Practice and Theory in Public Key Cryptography (PKC), 2010.