International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Privacy Preserving Multi-Keyword Graded Search on Encrypted Cloud Data with Integrity Checking

Snehal M. Shewale[1], Prof. Y. B. Gurav[2]

[1]Svitribai Phule Pune University, Department of Computer Engineering,
Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

[2]Professor, Svitribai Phule Pune University, Department of Computer Engineering,
Padmabhooshan Vasantdada Patil Institute of Technology, Pune, India

Abstract: ***Many of data owner use the public cloud to outsource their sensitive data and use data as they needed. In traditional method data is stored on local site or local server but it is not feasible for complex and sensitive data. Cloud computing allow data owner to store their data on remote site to reduce burden on local complex data storing. Data owners outsource their data to public cloud for safety and economic savings. When data owner outsource their data, this data is encrypted before it is transferred to the public cloud. In this paper we provide secure and privacy preserving technique for outsourcing data owner's sensitive data and retrieve relevance result for search on data. Here we use blowfish encryption algorithm for encryption of data to be outsourced. After the encryption of data it is stored on cloud server for further use. User or cloud service client need different kind of data. Cloud service user requests cloud data by multiple keyword query which is called as Multi keyword ranked search over encrypted data i.e. MRSE. The user queries are transferred to the cloud server. Server searches the relevant content by using the coordinate matching and sends the relevant results to the user. Data received from cloud server is in the encrypted format. Data owner provides access control to the user with key for decryption of data. Message Authentication Code algorithm is used to check and verify integrity of data. In this way this paper describes the technique of providing security to outsource data on cloud and checking the integrity of data.***

Keywords: Cloud server, Multi-keyword, Data Owner, Encryption, Integrity, Privacy.

1. Introduction

Cloud computing is defined as a type of computing that based on sharing computing resources instead of having local servers or personal devices to handle applications. Cloud computing has main three types Public Cloud, Private Cloud, Hybrid Cloud. Public Cloud provides services that are open for public use. Public cloud services may be available free. Private Cloud is cloud infrastructure used for single organization and managed internally. Hybrid cloud is combination of two or more cloud that may different entities but bound together. Cloud computing is concerns about privacy because the service provider can access the data that is on the cloud at any time. It could accidentally or deliberately alter or even delete information. Many cloud providers can share information with third parties . That is permitted in their privacy policies which users have to agree to before they start using cloud services. Users can encrypt data that is processed or stored within the cloud to prevent unauthorized access.

Cloud computing is an effective approach to deal with big data, through providing on-demand high quality services from powerful and configurable computing resources. One of the application of cloud computing is database as a service"[1]. This provides management of data on cloud. To protect data privacy against attacks from the cloud server, confidential data must be encrypted before being uploaded to the cloud server. This makes it difficult to perform traditional query processing operations. while outsourcing data there is problem of the confidentiality requires hiding the values and the relative order of the attribute in records from the cloud server, but computing a range query requires comparing the values of this attribute.

This paper proposes an efficient indexing method to support faster query evaluation than the trivial linear scan manner. On the one hand, to achieve the effective data retrieval need, the large amount of documents demand the cloud server to perform result relevance ranking, instead of returning undifferentiated results. Such ranked search system enables data users to find the most relevant information quickly [10]. Ranked search can also reduce unnecessary network traffic by sending result of query as only the most relevant data, which is highly desirable in the pay-as-you-use cloud paradigm. When we consider privacy protection, this ranking operation, however, should not leak any keyword related information. On the other hand, to improve the search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keywords search, as single keyword search often yields far too coarse results [1]. Coordinate matching,[11] i.e., as many matches as possible, is an efficient similarity measure among such multi keyword semantics to refine the result relevance, and used in the plaintext information retrieval (IR) community[1].Integrity of data check for accuracy of data .

The remainder of this paper is organized as follows: section II presents the related work with respect to the encryption of data and search on data stored in public cloud. Section III describes architecture for the system and modules to implements the system also describes proposed system algorithm ,Blowfish algorithm for encryption of data , followed by Section IV, which describes experimental results of implementation. Section V presents conclusion of the paper.

2. Related Work

In cloud computing, several existing schemes that had been attempted to implement data privacy & security, data privacy and access control. Data is stored as public or private so different searching strategies are available for both types of data. The confidential data are stored in the cloud in encrypted format so only the authenticated members who know the key can access the data. Y.C. Chang and M.Mitzenmacher pointed the problem of efficiently retrieve some of the encrypted files containing specific keywords, keeping the keywords themselves secret and not to endanger the security of the remotely stored files [2]. In solutions for this problem under well-defined security requirements are offered. This method is efficient because of no public key cryptography is used. M. Bellare A. Boldyreva describes Deterministic and Efficiently Searchable Encryption where the encryption algorithm is deterministic. Consider application of outsourced databases, where data is sent to a remote server. The database server is untrusted [3]. The data in each field in the database is encrypted separately under the public key of a receiver, who needs to be able to query the server to retrieve the encrypted records containing particular data. Deterministic encryption provides a possible solution to the problem [3]. J. Li, Q. Wang,C. Wang, N. Cao, K. Ren, and W. Lou,[4] introduces the fuzzy keyword search over encrypted cloud data and also maintaining the keyword privacy. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou,[5]highlight the challenge of secure ranked keyword search over encrypted cloud data also Ranked search greatly enhances system usability by enabling search result relevance ranking instead of sending undifferentiated results, and further ensures the file retrieval accuracy.

3. Proposed System

In this section, we present system architecture for proposed system.Figure.1 depicts architecture of proposed system for isolation preserved multi-keyword search over encrypted cloud.

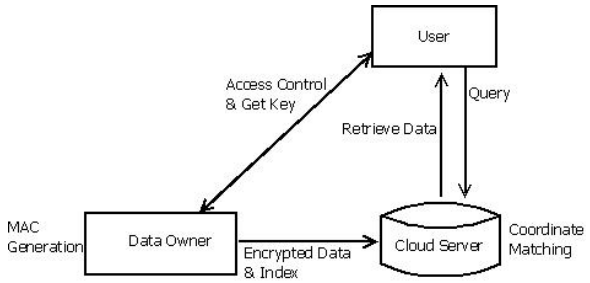

Figure 1: System Architecture

The system categorized into three entities Data Owner, Cloud Server, User or client. Data owner contains collection of data documents outsource on cloud server in encrypted form. Before outsourcing encrypted index for document is build. Outsourcing both index and encrypted document collection to Cloud Server. Cloud Server contains encrypted documents and index outsourced by the Data owner. Client provide query for searching document. Data integrity is check by Message authentication Code Algorithm.

The system is divided into following modules:
1. Binary data generation Module
2. Data ciphering Module
3. Data user access control Module
4. Data user query Module

In Binary data generation Module Data owner select the data and create the bit vector for that data. Using that bit vector of the data the binary data is generated. The binary data is the index for the data in the data owner. The bit vector is the bytes form of the data in the data owner. The bit vector is converted into the binary data. This bit vector and the binary data are ready for the data ciphering. Before that the message authentication code is generated for the data.

In Data ciphering Module the Data owner have to encrypt the original data by blowfish algorithm and send it to server. Then encrypt the binary data or the index and send it to server. Service provider did not know about the original content in the data owner. These indexes are used to refer the data in the service provider. It gives more security in the server side, so that the attackers can't use the data. Our system must prevent Server from learning any additional correspondence between plaintext values and cipher-text values except those obtained by prior knowledge. That is, we must protect the plain-text values for any encrypted records or queries from being disclosed to Server.

In Data user access control Module the user needs data from the cloud server. The users have different choices and the user send the query to the server or service provider. Before that the user gets the access from the data owner. For that the users end the details about him or her to the data owner. Then only the data owner receives the information from client and ready to send the decryption key. the access control mechanism is employed to manage decryption capabilities given to users This is a distributed setting where Server is on the remote side and not trusted. In Data user query Module the data user query is processed by the service provider. The service provider generates the bit vector for the query from the client. Then the service provider converts the bit vector into binary data. Service provider finds the similar data from the index. And send the encrypted data to the client. Then the client decrypts the received data by the key from the data owner using blowfish algorithm. And checks the integrity of the data by using the message authentication
Code

3.1 Proposed system Algorithm

Step 1: Authentication for data owner.
Step2: Data owner select data for Binary data generation and Message Authentication code generation
Step 3: Data owner have to encrypt the original data by blowfish algorithm and send it to server. Then encrypt the binary data or the index and send it to server.
Step 4: Provide access from the data owner to data user for data decryption.
Step 5: Data user checks the integrity of the data by using the Message authentication code Algorithm.

3.2 Blowfish Encryption Algorithm

Blowfish is a symmetric block cipher that can be effectively used for encryption and safeguarding of data. It takes a variable-length key, from 32 bits to 448 bits, making it ideal for securing data.

Blowfish Encryption Algorithm is encryption technique for transforming plaintext data into cipher text, this algorithm generate key for encryption .Input to this algorithm is document selected by the data owner for outsourcing .output is cipher text of data in document.

- Key Generation:
  Blowfish uses a large number of sub keys. The P-array consists of
  18, 32-bit sub keys: P1, P2,...,P18
- Encryption:

- Blowfish has 16 rounds.
- The input is a 64-bit data element, x.
- Divide x into two 32-bit halves: xL, xR.
- Then, for i = 1 to 16:
xL = xL XOR Pi
xR = F(xL) XOR xR
Swap xL and xR
- After the sixteenth round, swap xL and xR again to undo the last swap.
- Then, xR = xR XOR P17 and
xL= xL XOR P18.
- Finally, recombine xL and xR to get the cipher text.

- Decryption:

-Decryption is exactly the same as encryption, except that P1, P2,..., P18 are used in the
reverse order.

4. Experimental Result

In this proposed architecture main goal is to provide efficiency and privacy of data. System also verifies data integrity of data.

To get result we choose data set e.g. book dataset for proposed system .Bit vector and Binary data generation for selected data provides privacy to data. Blowfish Encryption Algorithm provides more security to the outsourcing data compare to previous encryption technique for data Outsourcing. This prevent cloud server from learning additional information.

Only user with authentication by Data owner can perform search on cloud data. Cloud perform relative ranking for search query. Relevance result of search query is received to user.i.e provide a effective Data retrieval. A multi-keyword search method must provide the following user and data privacy properties

1) (Data Privacy) No one but the user can learn the actual retrieved data.
2) (Index Privacy) The search index or the query index do not leak any information about the corresponding keywords.
3) (Trapdoor Privacy) Given one trapdoor for a set of keywords, the server cannot generate another valid trapdoor.
4) (Non-Impersonation) No one can impersonate a legitimate user.

Efficiency of Data Retrieval as show in figure 2

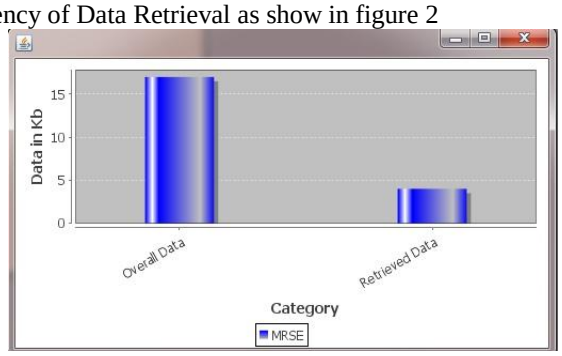

Figure 2 : Efficiency of Data Retrieval

Timing efficiency comparison for proposed system algorithm and existing system is as shown in figure 3

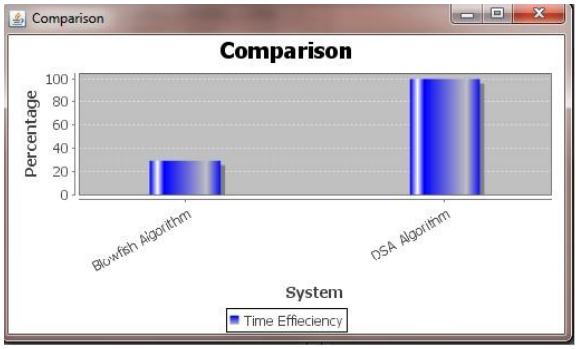

Figure 3: Timing Efficiency

5. Conclusion

In this paper, we proposed technique for efficient multi keyword search on encrypted cloud data. Bit vector and Binary Data generated for preserving data privacy. This system provides efficiency as receiving relevance result to user instead of differential result .This technique provides privacy in terms of keyword privacy, data privacy, Index privacy. Data Integrity checking is performed using Message Authentication Code Algorithm.

References

[1] N.Cao,C.Wang,M.Li,K Ren and Lou, "Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data",IEEE Transactions on Parallel & Distributed System Vol.25, January2014.
[2] Y.C. Chang and M. Mitzenmacher, "Privacy Preserving Keyword Searches on Remote Encrypted Data", Proc. Third Intl Conf.Applied Cryptography and Network Security, 2005.
[3] M. Bellare, A. Boldyreva, and A. O Neill," Deterministic and Efficiently Searchable Encryption", Proc. 27th Ann. Intl Cryptology Conf. Advances in Cryptology (CRYPTO 07), 2007.

[4] [4] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou," Fuzzy Keyword Search Over Encrypted Data in Cloud Computing", Proc. 27th Ann. Intl Cryptology Conf. Advances in Cryptology (CRYPTO 07), 2007.

[5]] N. Cao, C. Wang, M. Li, K.Ren, and W. Lou, "Enabling Secure and Efficient Ranked keyword Search over Outsourced Cloud Data",IEEE Transactions on Parallel & Distributed System Vol.23 No.8, August,2012.

[6] J. Goh, Secure Indexes, Cryptology ePrint Archive,http://eprint.iacr.org/2003/216

[7] N. Cao, S. Yu, Z. Yang, W. Lou, and Y. Hou," LT Codes-Based Secure and Reliable Cloud Storage Service",IEEE Transactions on Parallel & Distributed System 2012.

[8] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A Break in the Clouds: Towards a Cloud Definition" ,ACM SIGCOMM Computer. Commun. Rev., vol. 39, no. 1, pp. 50-55 2009 .

[9] M. Li, S. Yu, N. Cao, and W. Lou, "Authorized Private Keyword Search over Encrypted Data in Cloud Computing" ,Proc. 31st Intl Conf.Distributed Computing Systems (ICDCS 10), pp. 383- 392, June 2011.

[10] S. Kamara and K. Lauter," Cryptographic Cloud Storage" ,Proc. 14th Intl Conf. Financial Cryptography and Data Security, Jan. 2010.

[11] I.H. Witten, A. Moffat, and T.C. Bell, "Managing Gigabytes: Compressing and Indexing Documents and Images". ,Morgan Kaufmann Publishing,May 1999.