International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

Secure and Effectual Cloud Data Deduplication Verification

[1]Lingeshwaran. B, [2]Tamilselvan. V, [3]Vigneshwaran. V

Department of Computer Science and Engineering,
Sri Muthukumaran Institute of Technology, Chennai, Tamil Nadu, India

ABSTRACT

Secure search techniques over encrypted cloud data allow an authorized user to query data files of interest by submitting encrypted query keywords to the cloud server in a privacy-preserving manner. However, in practice, the returned query results may be incorrect or incomplete in the dishonest cloud environment. For example, the cloud server may intentionally omit some qualified results to save computational resources and communication overhead. Thus, a well-functioning secure query system should provide a query results verification mechanism that allows the data user to verify results. In this paper, we design a secure, easily integrated, and fine-grained query results verification mechanism, by which, given an encrypted query results set, the query user not only can verify the correctness of each data file in the set but also can further check how many or which qualified data files are not returned if the set is incomplete before decryption. The verification scheme is loose-coupling to concrete secure search techniques and can be very easily integrated into any secure query scheme. We achieve the goal by constructing secure verification object for encrypted cloud data. Furthermore, a short signature technique with extremely small storage cost is proposed to guarantee the authenticity of verification object and a verification object request technique is presented to allow the query user to securely obtain the desired verification object. Performance evaluation shows that the proposed schemes are practical and efficient.

1. INTRODUCTION:

The secure keyword search issues in cloud computing have been adequately researched. The issue in cloud security aims to continually improve search efficiency, reduce communication and computation cost, and enrich the search function with better security privacy protection. A common basic assumption of all these schemes is that the cloud is considered to be an ”honest-but-curious” entity as well as always keeps robust and secure software/hardware 3 environments. As a result, under the ideal assumption, the correct and complete query results always are unexceptionally returned from the cloud server when a query ends every time. The secure search is a technique that allows an authorized data user to search over the data owner’s encrypted data by submitting encrypted query keywords in a privacy-preserving manner and is an effective extension of traditional searchable encryption to adapt for the cloud computing environment. It is motivated by the effective information retrieve on encrypted outsourced cloud data.

2. EXISTING SCENARIO:

A matter of public concern is how to guarantee the security of data that is outsourced to a remote cloud server and breaks away from the direct control of data owners. Encryption on private data before outsourcing is an effective measure to protect data confidentiality. However, encrypted data make Effective data retrieval a very challenging task.

First introduced the concept of searchable encryption and proposed a practical technique that allows users to search over encrypted data through encrypted query keywords.

Later, many searchable encryption schemes that are necessary for its enhancement were proposed based on symmetric key and public-key setting to strengthen security and improve query efficiency with the growing popularity of cloud computing, how to securely, effectively and efficiently search over encrypted cloud data becomes a research focus.
Some approaches have been proposed based on traditional searchable encryption schemes. This existing system has various drawbacks as follows:

- Cloud server may return erroneous or incomplete query results once he behaves dishonestly for illegal profits.
- A latent space with lower-dimensionality while preserving important discriminative features amongst users.
- To learn an effective latent representation, we simultaneously incorporate prior knowledge, such as temporality of wellness features and heterogeneity of users.
- We first present the notations and then formally define the problem of representation learning of longitudinal data.

3. PROPOSED ALGORITHM:

AES Algorithm:

AES (acronym of Advanced Encryption Standard) is a symmetric encryption algorithm. The algorithm was developed by two Belgian cryptographer Joan Daemen and Vincent Rijmen. AES was designed to be efficient in both hardware and software, and supports a block length of 128 bits and key lengths of 128, 192, and 256 bits.

High speed and low RAM requirements were criteria of the AES selection process. As the chosen algorithm, AES performed well on a wide variety of hardware, from 8-bit smart cards to high-performance computers.

The features of AES are as follows:

- Symmetric key and a symmetric block cipher
- Stronger and faster than Triple-DES
- Provides full specification and design details
- Software implementable in C and Java

4. SYSTEM ARCHITECTURE:

A secure and fine-grained query results verification scheme by constructing the verification object for encrypted outsourced data files.
When a query ends, the query results set along with the corresponding verification object are returned together, by which the query user can accurately verify:

1) The correctness of each encrypted data file in the results set
2) How many qualified data files are not returned and
3) Which qualified data files are not returned.

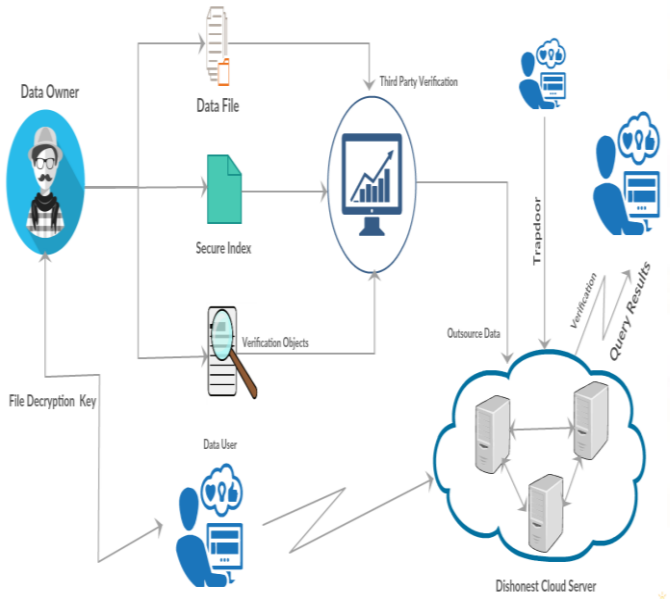

Figure 1: Block diagrammatic representation

Furthermore, our proposed verification scheme is lightweight and loose-coupling to concrete secure query schemes and can be very easily equipped into any secure query scheme for cloud computing. Just as possibly tampering or deleting query results, the dishonest cloud server may also tamper or forget verification objects themselves to make the data user impossible to perform verification operation. Specially, once the cloud server knows that the query results verification scheme is provided in the secure search system, this information may leak query user's privacy and expose some useful contents about data files. More importantly, this exposed information may become temptations of misbehavior for the cloud server.

6. MODULE IMPLEMENTATION:

6.1. DATA OWNERS
6.2. OWNER DATASET
6.3. THIRD PARTY VERIFIER
6.4. SHARED DATASET
6.5. SECURITY

MODULES DESCRIPTION:

6.1. DATA OWNERS:

Data owner can upload data's, that data are split into part data then send to trusted data checker, job of the data checker is to generate signature key from MD5 and compare with previous keys, if mismatch then that data send to Key generator Server , Job of the key generator are generate encryption key as user specified algorithm ,finally encrypt then store in Database.

6.2. OWNER DATASET:

In this Module it creates a data owner dataset, this dataset only map owner with our upload data's , we maintain common database for effectively find duplications. The files will be uploading only once. If another data owner going to upload the same file in database means they will get the notification (the data is already uploaded in database).So data owner can save cost and time.

6.3. THIRD PARTY VERIFIER:

In this modules, the third party auditor checks for the file integrity. If the file contains the same word as was in the file previously saved in the cloud then file will not store instead it shows error . The TPA will filter the file. If the file has some updation with uniqueness then TPD will accept the file and encrypt the file and stored to the cloud.

6.4. SHARED DATASET:

Share Dataset is an light weight dataset that only contain mapping file metadata information, in our project we maintain one common big data database instead of unique because efficiently find duplication and memory management, if data owner share our data to client that data not replicate instead map client name. Data deduplication enables data storage systems to find and remove duplication within data without compromising its availability.

6.5. SECURITY:

We are implementing "Dynamic Encryption key Generation". It means all shared data only view with data owner permission, so we can avoid from unknown access. Social users are group members they can only view and share the data. If want show the data mean they need to get permission to data owner then data owner will send Encryption key after they can view the data. If data owner does not provide the KEY mean user cannot view the file. data encryption provides an important guarantee for the security and privacy of clients' data, it limits the manners of the accessibility and availability of the encrypted data

7. SYSTEM CONFIGURATIONS AND IT'S IMPLEMENTATIONS:

7.1. HARDWARE SYSTEM CONFIGURATION:

Processor - Intel
Speed - 1.1 GHz
RAM - 2 Gb (min)
Hard Disk - 20 GB
Floppy Drive - 1.44 MB
Key Board - Standard Windows Keyboard
Mouse - Two or Three Button Mouse
Monitor - SVGA

7.2. SOFTWARE SYSTEM CONFIGURATION:

Operating System - Windows 7/8/10
Front End - HTML, J2EE
Scripts - JavaScript.
Server side Script - Java Server Pages.
Database - MySql
Database Connectivity - JDBC.

7.3. SOFTWARE DESCRIPTION:

7.3.1. Front End:

Java:

Java is a set of several computer software and specifications developed by Sun Microsystems, later acquired by Oracle Corporation, that provides a system for developing application software and deploying it in a cross-platform computing environment. Java is used in a wide variety of computing platforms from embedded devices and mobile phones to enterprise servers and supercomputers. While less common, Java applets run in secure, sand boxed environments to provide many features of native applications and can be embedded in HTML pages.

Writing in the Java programming language is the primary way to produce code that will be deployed as byte code in a Java Virtual Machine (JVM); byte code compilers are also available for other languages, including Ada, JavaScript, Python, and Ruby.

In addition, several languages have been designed to run natively on the JVM, including Scala, Clojure and Groovy. Memory management is handled through integrated automatic garbage collection performed by the JVM.

Netbeans:

NetBeans is a software development platform written in Java. The NetBeans Platform allows applications to be developed from a set of modular software components called modules. Applications based on the NetBeans Platform, including the NetBeans integrated development environment (IDE), can be extended by third party developers. The NetBeans IDE is primarily intended for development in Java, but also supports other languages, in particular PHP, C/C++ and HTML5.NetBeans is cross-platform and runs on Microsoft Windows, Mac OS X, Linux, Solaris and other platforms supporting a compatible JVM. Recently, NetBeans IDE 8.0 was released on March 18, 2014.NetBeans has a roadmap document for release plans.

7.3.2. Back End:

MySQL, the most popular Open Source SQL database management system, is developed, distributed, and supported by Oracle Corporation.

- **MySQL is a database management system:**

A database is a structured collection of data. It may be anything from a simple shopping list to a picture gallery or the vast amounts of information in a corporate network. To add, access, and process data stored in a computer database, you need a database management system such as MySQL Server. Since computers are very good at handling large amounts of data, database management systems play a central role in computing, as standalone utilities, or as parts of other applications.

- **MySQL databases are relational:**

A relational database stores data in separate tables rather than putting all the data in one big storeroom. The database structures are organized into physical files optimized for speed. The logical model, with objects such as databases, tables, views, rows, and columns, offers a flexible programming environment. The SQL part of "MySQL" stands for "Structured Query Language". SQL is the most common standardized language used to access databases.

Depending on your programming environment, you might enter SQL directly , embed SQL statements into code written in another language, or use a language-specific API that hides the SQL syntax.

- **MySQL software is Open Source:**

Open Source means that it is possible for anyone to use and modify the software. Anybody can download the MySQL software from the Internet and use it without paying anything.

- **The MySQL Database Server is very fast, reliable, scalable, and easy to use:**

If that is what you are looking for, you should give it a try. MySQL Server can run comfortably on a desktop or laptop, alongside your other applications, web servers, and so on, requiring little or no attention. If you dedicate an entire machine to MySQL, you can adjust the settings to take advantage of all the memory, CPU power, and I/O capacity available. MySQL can also scale up to clusters of machines, networked together.

MySQL Server was originally developed to handle large databases much faster than existing solutions and has been successfully used in highly demanding production environments for several years. Although under constant development, MySQL Server today offers a rich and useful set of functions. Its connectivity, speed, and security make MySQL Server highly suited for accessing databases on the Internet.

- **MySQL Server works in client/server or embedded systems:**

The MySQL Database Software is a client/server system that consists of a multi-threaded SQL server that supports different backends, several different client programs and libraries, administrative tools, and a wide range of application programming interfaces (APIs).

- **A large amount of contributed MySQL software is available:**

MySQL Server has a practical set of features developed in close cooperation with our users. It is very likely that your favorite application or language supports the MySQL Database Server. A wizard based installation is proceeded. My SQL open source

software is provided under the GPL Licence. There are resources of software that contributes MySQL where it is available to a great extent.

8. MERITS OF THE PROPOSED SYSTEM:

- We formally propose the verifiable secure search system model and threat model and design a fine-grained query results verification scheme for secure keyword search over encrypted cloud data.
- We propose a short signature technique based on certificate less public-key cryptography to guarantee the authenticity of the verification objects themselves.
- We design a novel verification object request technique based on Parlier Encryption, where the Cloud server knows nothing about what the data user is requesting for and which verification objects are returned to the user.
- We provide the formal security definition and proof and conduct extensive performance experiments to evaluate the accuracy and efficiency of our proposed scheme.

9. FUTURE SCOPE:

The physical security of data centers is also important, just like reliable encryption of information. In future works, the minimum requirements for the current SSL protocol will be seriously changed. Due to constantly increasing security requirements, the physical access to the data center will also be severely limited, and to enter the protected premises the user will need not only an electronic key, but also a procedure for biometric scanning. This enhances the way of privacy one needs as essential in their life.

10. CONCLUSION:

We propose a secure, easily integrated, and fine-grained query results verification scheme for secure search over encrypted cloud data. Different from previous works, our scheme can verify the correctness of each encrypted query result or further accurately find out how many or which qualified data files are returned by the dishonest cloud server. A short signature technique is designed to guarantee the authenticity of verification object itself. Moreover, we design a secure verification object request technique, by which the cloud server knows nothing about which verification object is requested by the data user and actually returned by the cloud server. Performance and accuracy experiments demonstrate the validity and efficiency of our proposed scheme.

Cloud computing is viewed as the next generation architecture of IT companies. As promising as it is, cloud computing also brings forth many new security issues when users outsource sensitive data to cloud servers. To keep sensitive users' data confidential against untrusted servers, existing solutions usually apply cryptographic methods. With data encryption, the same file will become different from each other, thus deduplication which is widely adopted by cloud storage service providers meets some challenges. Current method to solve the problem is to make use of some information computed from the shared file to achieve deduplication of encrypted data, say convergent encryption. But this piece of information which is computable from the file via a deterministic public algorithm is not really meant to be secret. To this end, we propose a scheme to address the deduplication of encrypted data efficiently and securely with the help of ensuring the ownership of the shared file, encrypting data using keys at user's will and realizing the anonymous store through the digital credential.

REFERENCES:

1. D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. of IEEE S&P, 2000, pp. 44–55.
2. Y.-C. Chang and M. Mitzenmacher, “Privacy preserving keyword searches on remote encrypted data,” in Proc. of ACNS, 2005, pp. 391–421.
3. C. Wang, N. Cao, K. Ren, and W. Lou, “Enabling secure and efficient ranked keyword search over outsourced cloud data,” IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, 2012.
4. P. Golle, J. Staddon, and B. R. Waters, “Secure conjunctive keyword search over encrypted data,” in Proc. of ACNS, 2004, pp. 31–45.
5. D. Boneh and B. Waters, “Conjunctive, subset, and range queries on encrypted data,” in Proc. of TCC, 2007, pp. 535–554.
6. Y. Hwang and P. Lee, “Public key encryption with conjunctive keyword search and its extension to a multi-user system,” in Proc. of Pairing, 2007, pp. 2–22.
7. E. Shen, E. Shi, and B. Waters, “Predicate privacy in encryption systems,” in Proc. of TCC, 2009, pp. 457–473.

8. W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proc. of ACM SIGMOD, 2009, pp. 139–152.

9. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," in Proc. Of INFOCOM, 2011, pp. 829–837.

10. "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, 2014