Advance Searchable and Verifiable Data Protection for Cloud Data

Ms. Rupali H. Shende
Department Of Computer Engineering
R.M.D. SINHGAD SCHOOL OF ENGINEERING
Pune,India
rupalishende13@gmail.com

Asst. Prof. Pradnya Kasture
Department Of Computer Engineering
R.M.D. SINHGAD SCHOOL OF ENGINEERING
Pune,India
pradnyakasture.rmdssoe@sinhgad.edu

***Abstract*—Outsourcing data to a third-party administrative control, as is completed in cloud computing, it gives rise to security concerns. The data temper may occur due to attacks by other users and nodes within the cloud. Therefore, high-security measures are required to protect the data within the cloud. Cloud storage can allow users to store and access their files at any time, from anywhere, with any device. To ensure data security in subcontracting, the user of the data should periodically verify the integrity of the data. This document proposes to reliable and verifiable data protection scheme, supported by third parties, that uses cloud computing technology. For a better description of the protocol, first, this system presents a differentiated system model by the user and a cube data storage structure. Based on the new system model and data structure, the schema helps users verify the integrity of data uploaded or downloaded at any time and search academic data online with encrypted keywords. Data de-duplication is performed before uploading file on the cloud by MD5 algorithm so it reduces the storage space.**

***Index Terms*—Cloud storage, hashchain, lightweight cryptography, Symmetric key, Searchable encryption, Security,**

I. INTRODUCTION

The development of society is inseparable from scientific and technological progress, both of which must rely on theoretical innovation and upgrades. Scientific research from all fields in Volumes all aspects of people's lives. With the continuous development of research in various fields and the emergence of new fields, the achievements of various fields are becoming increasingly abundant. There are many reasons for the increase of scholarly data, including an increase in the number of scholars, the complexity of the networks of scholars, the diversification of magazines and journals, the growing readership and the continuous expansion of professional fields. People care about the development of scholarly big data because these data are related to the quality of life in the coming decades or even these schemes focus on protecting cloud storage data from different aspects. The cloud can be utilized to provide secure data storage, fast search services and data integrity verification for scholarly big data. Thus, a cloud-aided secure and efficient scholarly data application system that can reasonably store scholarly data needs to be proposed. There are many reasons for the increase of data, including an increase in the number of academics, the complexity of academic networks, the diversification of magazines and journals, the growing number of readers and the continuous expansion of professional sectors. People are concerned about the development of large academic data because these data are related to the quality of life in the coming decades or even these schemes are focused on protecting data storage in the cloud from several aspects. The cloud can be used to provide secure data storage, fast search services and data integrity verification for big data. Therefore, it is necessary to propose a data application system based on a secure and efficient cloud that can reasonably archive data.

A. MOTIVATION

On web large numbers of documents are stored in a cloud server, searching against a keyword will result into large number of documents, not related to topic. The formation of scholarly big data is actually the result of the great development of scientific theory research. Scholarly big data include research scholar's personal information, papers, experimental data sets, and results. These data may include information regarding the author's privacy and social relationships, copyright and right of authorship, and experimental data related to personal privacy, such as medical data. These data are complex and extremely important. If the coauthor of an academic achievement is likely to be tampered with, the author's academic reputation may be affected. If malicious users are using legitimate users identities to upload data to the system, the researcher's results may be tampered with or replaced.

B. OBJECTIVE

To provide data security and verification of cloud store data. To improve user searching on large cloud-stored data.

II. REVIEW OF LITERATURE

1. Algorithms play an important role in solving research problems. Scientific publications contain a large number of high-quality algorithms developed by professional researchers. In digital libraries, the possibility of extracting and cataloging these algorithms would introduce a series of interesting applications that include search, discovery, and analysis of algorithms. This system talked about the prototype Algorithm

See, a system for searching algorithms in large-scale academic documents, as well as an illustration of the real demonstration system [11].

2. With the rapid growth of digital publishing, the collection, management and analysis of academic information has become increasingly difficult. The term Big Scholarly Data was coined to obtain rapidly growing academic data, which contains information that includes millions of authors, documents, citations, figures, tables, academic networks, and digital libraries. Nowadays, several academic data can be easily accessed and powerful data analysis technologies are being developed that allow us to examine science itself with a new perspective. In this article, this system examines the background and state of the art of large academic data. First, This system introduces the antecedents of academic data management and relevant technologies. Second, This system examines methods of data analysis, such as statistical analysis, social network analysis, and content analysis to deal with large academic data. Finally, this system examines representative research topics in this area, including scientific impact assessment, academic recommendation, and expert research. For each problem, the antecedents, the main challenges, and the latest investigations are treated. These discussions are intended to provide a complete review of this emerging area. This survey document concludes with a discussion on open issues and promising future directions [12].

3. Propose a new algorithm of safe subcontracting for the exponentiation (variable-exponential, based on a variable) form the first place in the two templates of unreliable programs. In comparison with the advanced algorithm, the proposed algorithm is superior both in terms of efficiency and control. Based on this algorithm, This system shows how to obtain encryptions and secure signatures from Cramer-Shoup in outsourcing [13].

4. The proposed algorithm is more efficient than existing algorithms. On the other hand, this document points out that the first secure algorithm for simultaneous modular extinction proposed recently is not secure, in which confidential information can be disclosed to malicious servers. As a result, This system proposes a new and more efficient algorithm for simultaneous modular exponents. This system also offers constructions to encrypt Cramer-Shoup and secure outsourcing firms that are also more efficient than the latest generation algorithms [14].

5. It presents the system that formalizes the notion of a verifiable database with incremental updates. In addition, This system proposes a general Inc-VDB structure that incorporates the primitive of vector commitment and the cryptographic and then incremental mode of MAC encryption. This system also presents a concrete Inc-VDB scheme based on the computational hypothesis of Diffie-Hellman. In addition, This system demonstrates that this system construction can achieve the desired safety properties[15].

6. Transformation of an encryption scheme based on anonymous identity (IBE) into a secure PEKS scheme that, unlike the previous one, guarantees consistency. Finally, This system suggest three extensions of the basic concepts discussed here, namely, anonymous cryptography based on hierarchical identity, public key cryptography with temporal keyword research, and identity-based cryptography with keyword research [1].

7. One of these system constructions, called RSA-DOAEP, has the additional feature of preserving length, so it is the first example of public key code. This system generalize this to obtain a notion of efficient search encryption schemes that allow more flexible privacy to the exchanges in terms of search through a technique called grouping[2].

8. Solve the challenging issue of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing. It displays ranked document according to matching result [4].

9. Performs searches with individual keywords and offers optimal server index sizes in an asymptotic manner, completely parallel search and minimal loss. this system implementation effort has led to several factors ignored by previous theoretical analyzes of coarse-grained performance, including the use of low-level space, I / O parallelism and smooth operation[5].

10. SSE scheme proposal to satisfy all the properties described above. this system construction expands the focus of the inverted Index in several non-trivial ways and introduces new techniques for designing SSE[8].

III. SYSTEM ARCHITECTURE/ SYSTEM OVERVIEW

The proposed system will provide security to data. Secure search protocol that can an efficient and easy-to-implement searchable and verifiable encryption scheme for search, In the proposed system, four entities are involved such as data owners, data users, cloud server and TPA. Data owners have a collection of files. Data owners upload the file then indexes will build. Data owners encrypt files and upload encrypted file to the cloud server. When data client wants to search over files from a cloud server, He enters a string to search. The system will give matched files. Then the client sends a request for the decryption key, the client will get that on mail. If key matches then the only file will download to the client. Then data client download files and decrypts these files. Third-party auditor checks the integrity of data and inform to the owner. In the proposed system contribute that data deduplication checking at the time of file uploading that

reduces storage space of storage server.

Live Survey

1. Banking Transaction.
2. Blockchain related application.
3. Healthcare related application in cloud storage.

A. Advantages

1. It provides searching in the way proposed string search on the cloud.
2. Provide integrity checking of data.

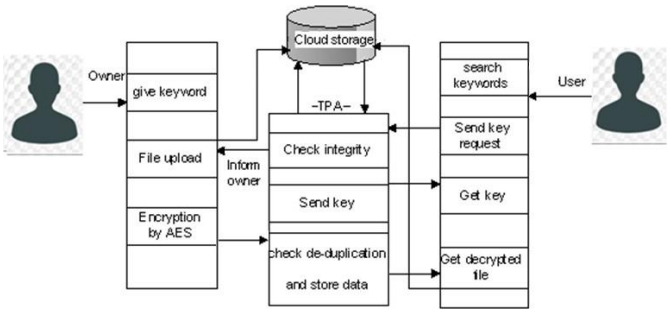

Fig. 1. Proposed System Architecture

B. Algorithms

1) MD5

The MD5 message digest algorithm is a most used hash function that produces a 128-bit hash value. Although the MD5 was originally designed to be used as a cryptographic hash function, it has been discovered that it suffers from extensive vulnerabilities. Still, It can be used as a checksum to verify the integrity of the data, but only against inadvertent corruption. It is suitable for other non-cryptographic purposes, for example, to determine the partition of a particular key in a partitioned database.

1. Data integrity check is the most common application of the hash functions. It is used to generate the checksums on data files.
2. Instead of storing the password in clear, mostly all login processes store the hash values of passwords in the file.

Steps:
Step 1-Append padded bits
Step 2-Append length
Step 3-Initialize MD Buffer
Step 4-Process message in 16-word blocks
Step 5-Output

2) Advanced encryption standard (AES) Algorithm For Encryption

It is symmetric algorithm. It used to convert plain text into cipher text.

Steps

Encryption

1: 128 bit data block
2: key expansion
3: add round key
4: sub byte, shift row, mix columns, add round key
5: sub byte, shift rows, add round key
6: 128 but encrypted block

Decryption

1: 128 but encrypted block
2: key expansion
3: add round keys, shift rows, subbytes
4: add round keys, mix columns, shift rows, sub bytes
5: add round key
6: 128 bit block

C. Mathematical Model

A block cipher is specified by an encryption function

$$E_K(P) := E(K, P) : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$$

which takes as input a key K of bit length k, called the key size, and a bit string P of length n, called the block size, and returns a string C of n bits. P is called the plaintext, and C is termed the ciphertext. For each K, the function EK(P) is required to be an invertible mapping on 0,1n. The inverse for E is defined as a function

$$E_K^{-1}(C) := D_K(C) = D(K, C) : \{0,1\}^k \times \{0,1\}^n \rightarrow \{0,1\}^n$$

taking a key K and a ciphertext C to return a plaintext value P, such that

$$\forall K : D_k(E_k(P)) = P all K : D_k(E_k(P)) = P.$$

D. HARDWARE AND SOFTWARE REQUIREMENTS

Hardware Requirements

1) Processor - Intel i5 core
2) Speed - 1.1 GHz
3) RAM - 2GB
4) Hard-Disk space - 40 GB
5) Key Board - Windows Keyboard
6) Mouse - Two or Three Button Mouse
7) Monitor - SVGA

Software Requirements

1) Operating System - XP, Windows7/8/10
2) Coding language - Java, MVC, JSP, HTML, CSS etc
3) Software - JDK1.7
4) Tool - Eclipse Luna
5) Server - Apache Tomcat 7.0
6) Database - MySQL 5.0

IV. SYSTEM ANALYSIS AND RESULT

A. Screen

Fig.2. Login

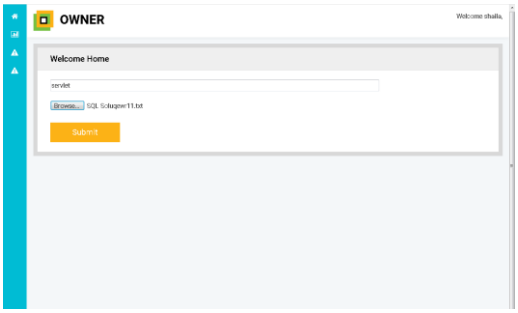

Fig.3. Select file

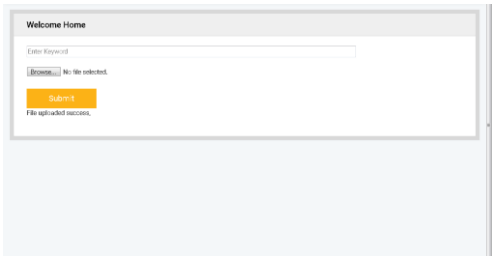

Fig.4. Upload file

B. System analysis

Experimental setup Table 1-

1.Setup: It consider This algorithm takes a security parameter k as input

2.Extract: It generate keys for each user file.

3.Upload: It consider to upload the file on cloud

4. Download: It consider time to download file from cloud

Fig.2 shows the pictorial representation of these required execution time in proposed system. X-axis is Phases and Y-Axis is time to execute in ms

The proposed system uses AES Algorithm that key consist some user attribute. In existing system key is single and no any attribute is considered. The uploading and downloading time is less because the algorithm used is symmetric encryption algorithm. That uses one private key that will used for both encryption and decryption of file.

TABLE I

EXECUTION TIME REQUIRED BY SETUP, EXTRACT, ENCRYPT AND DECRY PT ALGORITHMS IN DIFFERENT SCHEMES.

Index Number	System	Setup (ms)	Extract (ms)	To upload (ms)	To down-load (ms)
1	Identity Encryp-tion	3 0.311	0.645	10.88	1.422
2	Proposed system	1.458	0.640	4.340	1.156

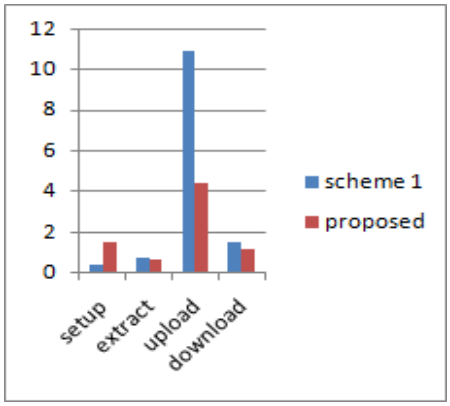

Fig. 5. Required Execution Time in Proposed System

V. CONCLUSION

Construct a system model that can distinguish the users according to their roles and special requirements of scholarly big data. Moreover, an innovative cube data storage structure is proposed. On the base of the novel system and data structure. This system proposed a real searchable and verifiable data protection scheme for scholarly big data. The security and performance analyses show that this system scheme is efficient for scholarly big data. Data de-duplication is performed.

REFERENCES

[1] Michel Abdalla, Mihir Bellare, Dario Catalano, Eike Kiltz, Tadayoshi Kohno, Tanja Lange, John Malone-Lee, Gregory Neven, Pascal Paillier, and Haixia Shi. Searchable Encryption Revisited: Consistency Properties, Relation to Anonymous IBE, and Extensions. volume 21, pages 350391. Springer, 2008.

[2] Mihir Bellare, Alexandra Boldyreva, and Adam ONeill. Deterministic and Efficiently Searchable Encryption. In Annual International Cryptology Conference, pages 535552. Springer, 2007.

[3] Dan Boneh, Giovanni Di Crescenzo, Rafail Ostrovsky, and Giuseppe Persiano. Public Key Encryption With Keyword

Search. In International Conference on the Theory and Applications of Cryptographic Techniques, pages 506522. Springer, 2004.

[4] Ning Cao, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. Privacy- Preserving Multi-Keyword Ranked Search Over Encrypted Cloud Data. volume 25, pages 222233. IEEE, 2014.
[5] David Cash, Joseph Jaeger, Stanislaw Jarecki, Charanjit S Jutla, Hugo Krawczyk, Marcel-Catalin Rosu, and Michael Steiner. Dynamic Search- able Encryption in Very-Large Databases: Data Structures and Implemen- tation. volume 2014, page 853. Citeseer, 2014.
[6] Yoshinao Uchide and Noboru Kunihiro. Searchable symmetric encryption capable of searching for an arbitrary string. Wiley Online Library, 2016.
[7] Reza Curtmola, Juan Garay, Seny Kamara, and Rafail Ostrovsky. Search- able Symmetric Encryption: Improved Definitions and Efficient Construc- tions. volume 19, pages 895934. IOS Press, 2011.
[8] Seny Kamara, Charalampos Papamanthou, and Tom Roeder. Dynamic Searchable Symmetric Encryption. In Proceedings of the 2012 ACM conference on Computer and communications security, pages 965976. ACM, 2012.
[9] Jonathan Katz and Yehuda Lindell. Introduction to Modern Cryptography. CRC press, 2014.
[10] Muhammad Naveed, Seny Kamara, and Charles V Wright. Inference Attacks on Property-Preserving Encrypted Databases. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pages 644655. ACM, 2015.
[11] S. Tuarob, S. Bhatia, P. Mitra, and C. L. Giles, Algorithmseer: A system for extracting and searching for algorithms in scholarly big data, IEEE Transactions on Big Data, Volume 2, no. 1, 2016.
[12] F. Xia, W. Wang, T. M. Bekele, and H. Liu, Big scholarly data: A survey, IEEE Transactions on Big Data, Volume 3, no. 1, 2017.
[13] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, New algorithms for secure outsourcing of modular exponentiations, IEEE Transactions on Parallel and Distributed Systems, Volume 25, no. 9, 2014.
[14] J. Shen, J. Shen, X. Chen, X. Huang, and W. Susilo, An efficient public auditing protocol with novel dynamic structure for cloud data, IEEE Transactions on Information Forensics and Security, Volume 12, no. 10, 2017.
[15] X. Chen, J. Li, J. Weng, J. Ma, and W. Lou, Verifiable computation over large database with incremental updates, IEEE transactions on Computers, Volume 65, no. 10, 2016.
cations Security (ASIACCS12), 2012, pp. 18.