A Secure Multi-Weighted Keyword Ranking Search on Encrypted Cloud Data

R. NallaKumar[1], K. Sivaranjani[2]
[1] Teaching Fellow, Department of computer science, Anna University Regional Office, Coimbatore, Tamil Nadu.
[2] PG research Scholar, Department of computer science, Anna University Regional Office, Coimbatore, Tamil Nadu.

Abstract – Cloud computing is one of the most increasing one with the increase no. of cloud users. In today's environment every user wants to store and access their data at any time and at anywhere. Cloud storage system allows users to access their stored data using several virtual machines. Due to this scalable nature, the clouds are vulnerable and possible for several security and privacy issues. In order to provide the privacy and security for the encrypted data, a new framework is proposed. The proposed framework is named as "PASS" (Privacy And Secure data Search"). The proposed PASS framework identifies security and privacy issues in cloud environment. The PASS framework helps to protect the user search privacy and content security over encrypted data. Compared with the existing schemes, the scheme only need to check a small portion of ranked indexes in a results and, thus, greatly reduces the verification cost. The PASS framework consist many algorithms and techniques, which are i) Hidden vector encryption algorithm, this contains the set of processes such as (Setup, Encrypt, KeyGen, indexing and verification) ii). Adaptive Agglomerative Clustering (AAC), this used to group the documents based on its similarity nature. iii) Dynamic Hash Tree with sporadic is used to update the indexing process. This will improve the search efficiency. iv) Bloom filter and bloom search for fast document search, this will adequately helps to handle huge number of clients in the common cloud environment. And the PASS scheme supports different multi-weighted keyword semantics over encrypted information and this also verifies the integrity of the order within the search result. The proposed system aims to achieve high security and privacy for cloud data with increased search efficiency, accuracy and time efficiency.

Index Terms – Cloud Computing, Data privacy, Hidden Vector Encryption, Clustering, Cloud service Provider, AAC, Security.

1. INTRODUCTION

Cloud computing is rising in terms of powerful resources and types of computing services in the last decade. It is a method of delivering technology to the consumer. Cloud computing provides active, scalable and pay-as-you-use service models. It can be adopted by many organizations to save their expenditure and time to build required IT infrastructure. The cloud computing is another process computing, distributed computing, parallel computing, virtualization technology, utility computing and other computer technologies. For example, large scale computation and data storage, virtualization, high expandability high reliability and low price service. The security issue of cloud computing is extremely paramount and it can keep the fast improvement of cloud computing. Cloud computing security is an advancing sub-area of network security. Using any device like PDA (Personal Digital Assistant), mobile etc, the cloud can be accessed through Cloud Service Provider (CSP). The cloud computing architecture consists of components of f platform, back end platform and a network host, for example PDA, mobile etc., computers, servers and data storage services. Cloud computing encompasses many technologies. Therefore security issues for many of these systems are applicable for cloud computing.

There are several security issues threatens the cloud environment by above issues. To preserve data privacy and oppose unsolicited accesses in the cloud and beyond it sensitive data, e-mails, personal health records, photo albums, tax documents and so on, may have to be encrypted by data owners before outsourcing to the commercial public cloud; this, however, obsoletes the traditional data utilization service based on plaintext keyword search. The inconsequential solution of downloading all the data and decrypting locally is clearly not viable due to the large amount of bandwidth cost incur in cloud environment. Exploring privacy preserving and effective search service over encrypted cloud data have great attention in the research. Considering potentially huge number of on demand data users and large amount of outsourced data documents in the cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability, and scalability.

In common scenario, the Document ranking is applied to yield fast and accurate search, however the priorities of all the documents is kept same so that the cloud service provider and third party remains unaware of the important documents and files while making the ranking, thus, maintaining privacy of data is challengeable. Ranked search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data, which is highly desirable in the "pay-as-you-use" cloud paradigm. For privacy protection, at the time ranking process the authority should not leak any keyword and its related information's. Moreover to improve search result accuracy as well as to enhance the user searching experience, it is also necessary for such ranking system to support multiple keyword searches, as single keyword search often yields far too coarse results. As a common practice indicated by today's web

search engines (ex. Google search), data users may tend to provide a set of keywords instead of only one as the indicator of their search interest to retrieve the most relevant data from the huge repository. With the consideration of data privacy and as well as effective data search, a real privacy should be performed. It is obtained only if the user's identity remains hidden from the Cloud Service Provider (CSP) as well as the third party user on the cloud server.

2. PROBLEM STATEMENT

Finding and analyzing relationship between documents and other digital data's are very tedious and it became very complicated when the data's are multimedia related content. That is more complex if the contents are encrypted or concealed. This type of search will lead to significant search accuracy performance degradation. Also the volume of data in data centers has experienced a dramatic growth. This will make it even more challenging to design encrypted search schemes that can provide efficient and reliable information search on large volume of encrypted content over cloud. But, applying the privacy and security on the data and user information in the encrypted cloud data search system remains a very challenging task because of inherent security and privacy obstacles, including various strict requirements like the data privacy, the index privacy, the keyword privacy, and many others.

We developed a new design and analyze a new Fast and Reliable privacy and security Techniques for effective data search on encrypted cloud data. The system also proposes a novel Key independent and fast & selective data indexing Technique for Confidentiality of the data search over the cloud by the end user. Deployment & Integration of different algorithms and techniques simultaneously for providing Confidentiality, Privacy, and Authentication of data over secure Cloud. The proposed system aims to reduce the verification cost and time over encrypted data. And this also aims to handle high volume of data, so the scalability will be considered. The system also has an objective to perform data dynamics and integrity in the private cloud.

3. RELATED WORKS

In [4] paper, authors proposed schemes in this paper support only Boolean keyword search. This scheme solves the problem of supporting efficient ranked keyword search over cloud data. Using this scenario, effective utilization of remotely stored encrypted data is achieved in Cloud Computing. Authors were mainly concerning on searching effective as well as secure ranked keyword searching for encrypted data. This system uses SSE technique for keyword searching. For ranking function TF*IDF rules were applied. For security purpose, OPSE crypto primitive is developed in this system.

In [5] authors define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) and they also concern with preserving strict system-wise privacy in the cloud computing paradigm. The proposed MRSE scheme aims to achieve various stringent privacy requirements in multiple and different threat methods. Coordinate matching technique is used to capture the relevance of data documents required for query. This system uses 'inner-product-similarity' to search number of keywords in the document. To attempt this purpose authors were proposing MRSE technique. Compare to other multi-keyword ranked searching technique this system produces very overheads.

In the paper [6], the author formalizes and solves the problem of effective fuzzy-keyword search on encrypted cloud data and maintains keyword privacy in every process. An advance technique is proposed in the paper. The technique is called as wild card based technique. This helps in searching fuzzy keywords. Fuzzy keyword search technique improves the usability of system by returning files with exactly matching keywords that are pre-defined. Proxy-server in this system is used to give response for receiver keyword query. PEKS does not have a requirement of coordination between sender and receiver when they are firstly join in opposite. This system requires special methods for sorting the keywords.

In this paper [7], authors proposed a probabilistic public key system namely, PEKS. This technique is more convenient to search cipher-texts for numerous users. This system achieves the multi-keyword search in fuzzy search. This system does not require any predefined keyword dictionary for keyword searching or keyword matching. This system adopts special hash function to build an index of searched keywords. The LSH function approach is used to build index as well as to provide secure fuzzy search in multi-keyword search.

In [8] paper, authors proposed a novel multi keyword fuzzy search scheme is for exploiting the locality-sensitive hashing technique. Instead of expanding the index file fuzzy matching is done through algorithmic design. This approach of leveraging LSH functions in the Bloom filter provides efficient solution to the secure fuzzy search of multiple keywords.

Literature summary:

In the literature, searchable encryption is a helpful technique that treats encrypted data as documents and allows a user to securely search through a single keyword and retrieve documents of interest. However, direct application of these approaches to the secure large scale cloud data utilization system would not be necessarily suitable, as they are developed as crypto primitives and cannot accommodate such high service-level requirements like system usability, user searching experience, and easy information discovery. Although some recent designs have been proposed to support Boolean keyword search as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality. The early work in the literature has been aware of this problem, and provided a solution to the secure

ranked search over encrypted data problem but only for queries consisting of a single keyword. How to design an efficient encrypted data search mechanism that supports multi-keyword semantics without privacy breaches still remains a challenging open problem. But, still effective techniques have to be identified for perfect data security process.

4. PROPOSED SYSTEM

The proposed work identifies security and privacy issues for secure data management in cloud environment. An efficient privacy preserving data search and verification scheme is proposed to protect the user search privacy and content security over encrypted data. Compared with the existing schemes, the scheme only need to check a small portion of ranked indexes in a results and, thus, greatly reduces the verification cost.

In the proposed system, explore supporting different multi-keyword semantics (e.g., weighted query) over encrypted information and checking the integrity of the order within the search result. In the proposed system, an adaptive agglomerative clustering (AAC) method is proposed to support more search semantics and also to meet the demand for fast cipher text search within a dynamic huge data environment. The proposed system aims to provide security and privacy for cloud data with increased search efficiency, accuracy and time efficiency.

Algorithm and techniques used in proposed system:

- Hidden vector encryption: this contains the set of processes such as (Setup, Encrypt, KeyGen, indexing and verification)
- adaptive agglomerative clustering (AAC) is proposed to group the documents based on its similarity nature.
- Dynamic Hash Tree with sporadic is used to update the indexing process. This will improve the search efficiency.
- Bloom filter and bloom search for fast document search, this will adequately helps to handle huge number of clients in the common cloud environment.
- Semantic calculation between terms is performed for data grouping and analysis.

Advantages of the proposed system:

- It reduces the verification cost and time.
- Handles high volume of data
- Performs multi-keyword search
- Provides fast search.
- Performs data dynamics and integrity in the private cloud

5. PERFORMANCE ANALYSIS

In analysis phase, the system analyzes the security, performance and experimental results PASS, which are described in the following sections. The results chapters prove the proposed system is outperformed than the existing techniques. This considered the verification delay and indexing key for deployed data on the cloud in the process of retrieval. Encryption and key generation and verification delay are specified below.

Encryption Process	**Data size(512 bits)**	**Data size(1024 bits)**
Existing System	8.4	14.9
Proposed system	3.6	6.7

Table 1.0 Encryption process with Data size

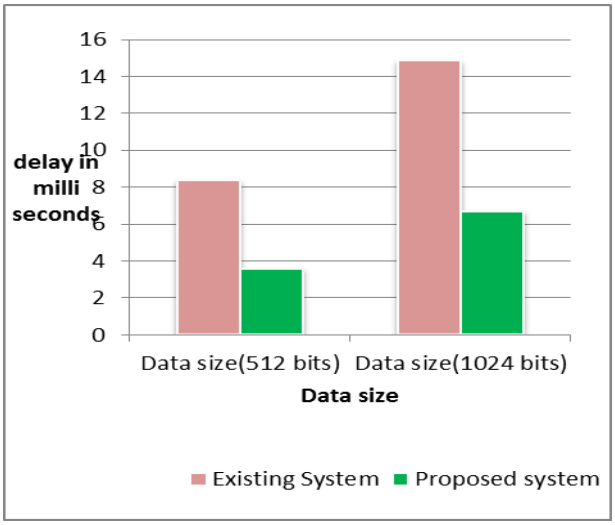

Figure 1.0 Delay comparison chart (encryption)

The above delay comparison chart indicates the execution time for the algorithm to produce cipher texts and corresponding keys before storing the data.

Verification and decryption delay are specified below.

Key verification Process	**Data size**	**data size**
Existing System	32.4	78.9
Proposed system	31.6	69.7

Table 2.0 Key Verification process

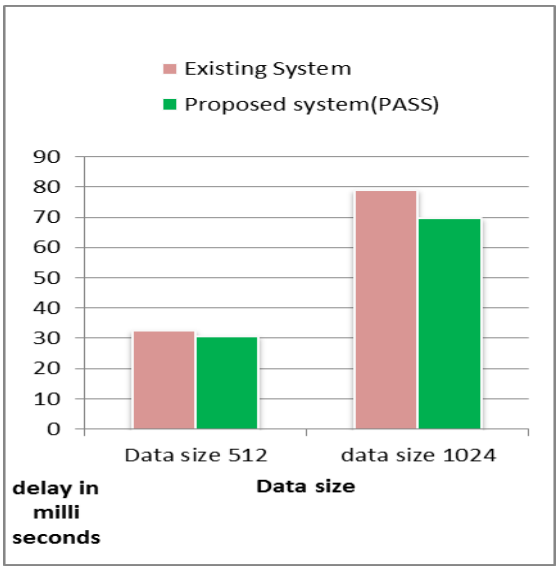

Fig 2.0 Delay comparison chart (Key verification)

The above delay comparison chart indicates the execution time for the algorithm to produce normal texts from the cipher data and corresponding keys before storing the data.

6. CONCLUSION

The proposed work mainly focused on providing privacy, security and integrity process to the data on cloud and the proposed system also performs multi-keyword ranked search over encrypted cloud data using efficient similarity measure of co-ordinate matching. The proposed system stringent privacy is provided by assigning the cloud user a unique ID. This user ID is kept hidden from the cloud service provider as well as the third party user in order to protect the user's data on cloud from the CSP and the third party user. Thus, by hiding the user's identity, the confidentiality of user's data is maintained. And this also uses an effective HVE algorithm to search the data on the encrypted content without affecting the security. With the help of different algorithms, the proposed system increases the privacy and security of the cloud data with reduced time and computational overhead.

REFERENCES

[1] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. IEEE Distributed Computing Systems (ICDCS'10), Genoa, Italy, Jun. 2010, pp. 253–262.

[2] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacypreserving multi-keyword ranked search over encrypted cloud data," in Proc. IEEE INFOCOM'11, Shanghai, China, Apr. 2011, pp. 829–837

[3] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. IEEE INFOCOM'10, San Diego, CA, Mar. 2010, pp. 1–5

[4] P. Xu, H. Jin, Q. Wu, and W. Wang, "Public-key encryption with fuzzy keyword search: A provably secure scheme under keyword guessing attack," Computers, IEEE Transactions on, vol. 62, no. 11, pp. 2266–2277, 2013.

[5] B. Wang, S. Yu, W. Lou, and Y. T. Hou, "Privacy-preserving multi-keyword fuzzy search over encrypted data in the cloud," in IEEE INFOCOM, Toronto, Canada, May 2014, pp. 2112–2120.

[6] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "A break in the clouds: towards a cloud definition," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 39, no. 1, pp. 50–55, 2009.

[7] S. Kamara and K. Lauter, "Cryptographic cloud storage," in *RLCPS, January 2010, LNCS. Springer, Heidelberg*.

[8] A. Singhal, "Modern information retrieval: A brief overview," *IEEE Data Engineering Bulletin*, vol. 24, no. 4, pp. 35–43, 2001.

[9] I. H. Witten, A. Moffat, and T. C. Bell, "Managing gigabytes: Compressing and indexing documents and images," Morgan Kaufmann Publishing, San Francisco, May 1999.

[10] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in *Proc. of S&P*, 2000.