Trends in Engineering and Technology (NCTET-2K17)
International Journal of Advanced Engineering Research and Science (IJAERS) *Special Issue-3*
https://dx.doi.org/10.22161/ijaers/nctet.2017.11 *ISSN: 2349-6495(P) | 2456-1908(O)*

Security Guaranteed High Performance Data Search on Encrypted Cloud Storages

B.Suresh Babu[1], S.Jeelan[2], V.Janardhan Babu[3],

[1]Asst.Professor, Department of CSE, SVPCET, PUTTUR.
Email: sureshbabu95@gmail.com [2]Assoc.Professor,
Department of CSE, SVPCET, PUTTUR. Email: jee.fuzi@gmail.com
[3]Professor, Department of CSE, SVPCET, PUTTUR.
Email: ungarala66@gmail.com,

Abstract*—The cloud computing environment provides storage and computing power to the users. The cloud data sharing scheme uses the encrypted cloud storage model to protect the data values. Encrypted cloud search is carried out on the indexed data values. High Performance Computing (HPC) models are adapted to improve the speed of the cloud search process. Graphical Processing Unit (GPU) architectures are also employed to improve the cloud search process. Statistical weight models are applied to estimate the score values for the term collections. Encrypted cloud data centers are constructed to provide privacy and security guaranteed data retrieval operations. Data retrieval is carried out with the supported of the Order Preserving Encryption (OPE) mechanism. The inverted index is performed with relevance score values. The OPE technique defends the score and index values. Encrypted data distribution is not considered in the deterministic OPE scheme. One to Many OPE or Probabilistic OPE scheme protects the encrypted data with its distribution levels. The data retrieval is performed with the binary search model. Differential attacks are initiated with the distribution interval analysis mechanism.The encrypted cloud search scheme is constructed with privacy and security features. The term subset reassignment model is adapted to defend the change point in the indexes. The index values are secured with noise keyword values. The data retrieval operations are designed with semantic relationship models. The semantic relationship based query model increases the accuracy with minimum computational overheads*.

Keywords — High Performance Computing, probabilistic schemes, Differential Attacks, Cloud Search

I. INTRODUCTION

Competitive business environments are putting pressure on IT managers to accomplish more each year with reduced budgets. With the need for flexibility, competitive edge and faster time to market, IT organizations must find new solutions that are more efficient and more cost-effective than their past or current solutions. The original data center started as a private server room hosted within the organization's facility containing many individual servers running single applications. In the early days of data centers, most organizations were responsible for maintaining the servers and software and required a number of personnel resources to manage the servers as well as the facility.

Some larger organizations continue to manage internal data center, many business managers are able to increase service levels, cover more users and lower response times by outsourcing their out-dated server farms to third-party data centers and cloud computing providers. These third-party data center providers are better equipped to maintain and update server equipment. This white the system defines data centers and explore cloud networking[11]. A data center is a centralized repository for the storage, management and dissemination of data and information. Typically, a data center is a facility used to house computer systems and associated components, such as telecommunications and storage systems. Often times, there are redundant or backup power supplies, redundant data communications connections, environmental controls and security devices[12].

One key benefit to the data center is that physical hard drive storage resources are aggregated into storage pools, from which "logical storage" is created. The heterogeneous nature of most storage systems allows many different vendors' storage hardware to be added to the system with little or no noticeable effect. These logical storage spaces can be reached from many different computer systems that share the same pool of storage space. One of the biggest benefits to storage virtualization

– other than the obvious ones such as centralized backups and the need for fewer hard drives overall – is the fact that the data can be replicated or migrated to another location transparently to the server using the logical storage point.

One of the not so glamorous or "hi-tech" benefits of the data center is the consolidation of all of the facility resources such as HVAC, electrical, network connections, wiring, hardware, software and personal. Many corporations have multiple server rooms with duplicated services across their entire organization, all of which are running on duplicated hardware and software platforms. In an attempt to reduce duplication and wasted expense, many corporations are consolidating their server rooms into private data centers, reducing the duplication of hardware, software and facilities needed to operate their business.

The cloud data sharing environment is build with a set of elements. They are data center, data provider and client environments. The data center provides the shared storage space to the data providers. The data provider stores the shared data values under the storage space allocated in data

https://dx.doi.org/10.22161/ijaers/nctet.2017.11 ISSN: 2349-6495(P) | 2456-1908(O)

centers. The data providers are controlled by multiple users. The client application is used to
download and access the shared data values. The data integrity is verified under the data provider or third party environments. The third party verification model is referred as public audit ability. The data provider based verification is denoted as private audit ability.

II. RELATED WORK

Cabarcos P.A. et al [1] in 2012 proposed a novel middleware architecture that allows sessions initiated from one device to be seamlessly transferred to a second one under a cloud computing environment. Díaz-Sánchez D. et al [2] presented a cloud computing middleware Media Cloud for Set-top boxes for classifying, searching, and delivering media inside home network and across the cloud. Seung Gwan Lee et al [3] proposed a personalized DTV Program Recommendation system under a cloud computing environment. The system can analyze and use the viewing pattern of consumers to personalize the program recommendations. Grzonkowski S. et al [7] proposed a user centric approach to authentication for home networks. This approach enables the sharing of personalized content and more sophisticated network-based services over a conventional TCP/IP infrastructure. Sanchez R. et al [8][11] proposed an IdM architecture based on privacy and reputation extensions to enable the global scalability and usability for consumer cloud computing paradigm. However, all these services are likely to be available to consumers only with the premise that an effective and efficient cloud search service is achieved.

To apply the searchable encryption to cloud computing, some researchers have been studying further on to search over encrypted cloud data efficiently. Li et al. [9] firstly proposed a fuzzy keyword search scheme over encrypted cloud data, which combines edit distance with wildcard-based technique to construct fuzzy keyword sets, to address problems of minor typos and format inconsistence. Wang et al. [10] proposed a ranked search scheme, in which by giving each keyword a weight TF-IDF, the cloud server can rank relevant data files with no knowledge of a specific keyword weight. But this scheme supports only single keyword search. Then Cao et al. [5] proposed a ranked scheme supporting multi-keyword, which uses a vector space model and characteristics of matrix to realize trapdoor penetrations levels and thereby preserves data privacy. Chai et al. [6] proposed a verifiable symmetric search encryption scheme, which can prove the correctness and completeness of results. Sun et al. [4] also proposed a multi-keyword ranked search scheme based on vector space model (VSM). The VSM can measure the similarity between document index vector and query vector and hence support more accurate ranked search results. But this scheme cannot support semantics-based search.

III. ENCRYPTED CLOUD DATA SEARCH

Companies and individuals need more and more storage space and computing power, that is why the cloud computing is increasingly used since its appearance. Cloud computing allows users to pay for what they use instead of buying their own servers which involves significant economic savings.

The data are often sensitive and confidential. Therefore, It is primordial to protect the outsourced data against possible external attacks and the cloud server itself. For this reason, the data must be encrypted before being outsourced. Users tend to take advantage of the large storage space offered by the cloud to store a huge number of documents. This can complicate the user's task to retrieve a specific document. To overcome this problem, the use of an information retrieval system (IRS) becomes necessary into a cloud server.

Considering that the data hosted in the cloud server are encrypted. The issue is that a traditional IRS works only on clear data. consequently, it is necessary to have an IRS able to search over encrypted data.

Many approaches have been proposed in the literature in order to perform search without any sensitive information leakage. The common point between them is the use of encrypted queries called trapdoors by the data users and a secure index to search over a collection of encrypted documents by the cloud server. After that, some works have tried to improve the quality of search by proposing semantic search approaches that exploit semantic graphs and ontology's order to improve the recall and the precision. For privacy and feasibility concerns, we use the vector representation of documents and queries which is the most used in the literature, instead of an inverted index. Unfortunately, this kind of representation has a negative effect on the search performance in term of processing time, given that, during the search process, the query vector is compared with each document vector. For example, if there are one million documents, the similarity function which is time consuming operation will be called one million times.

Very few studies have focused on the search performance by compressing the index in order to accelerate the search process. However, this technique causes the degradation of the quality of results without significantly improving the performance.

To the best of our knowledge, all prior works assume that the server receives only one query at the same time. In practice, the server may receive several queries sent from different users simultaneously making the search performance issue more complicated, given that the search process is already slow while treating a single query. To overcome this challenge we propose several techniques that allow the server to treat multiple queries simultaneously and respond to each user within a reasonable time.

The goal of this work is to propose some techniques that allow accelerating any searchable encryption scheme that uses a vector model to represent documents and queries without any degradation on the search quality in term of recall and precision. Our proposed parallelization techniques exploit several HPC architectures such as Multi-core CPU processor, Computer Cluster and Graphic Processing Unit (GPU) in order to accelerate the search process by distributing the work between several processes and treating lots of queries simultaneously. This solution allows achieving a speed-up of 46x.

https://dx.doi.org/10.22161/ijaers/nctet.2017.11 ISSN: 2349-6495(P) | 2456-1908(O)

The first challenge of our work is to find the most appropriate way to parallelize the search process and take benefit of each HPC architecture while respecting its programming model. The second one consists of reducing the response time of the server without degrading the search performance by treating several queries simultaneously.

IV. PROBLEM STATEMENT

Encrypted cloud storage is used to share user data with security and privacy. Raked search in encrypted cloud data process is carried out using Order Preserving Encryption (OPE) technique. Order Preserving Encryption (OPE) is applied to encrypt relevance scores of the inverted index. In deterministic OPE the cipher texts reveals the distribution of relevance scores. One-to-many OPE is employed to flatten the distribution Differential attack on one-to-many OPE is initiated by exploiting the differences of the ordered cipher texts. The following drawbacks are identified from the existing system. The following problems are identified from the existing system. Hardware dependant acceleration mechanism. Keyword inferring process is not controlled. Change point analysis based relevance score distribution estimation is not handled. Background knowledge based attacks are not controlled. Semantic query model is not provided

V. HIGH PERFORMANCE DATA SEARCH WITH SECURITY ON ENCRYPTED CLOUDS

The cloud computing environment provides high scalable data storages to the users. The data owner uploads the shared data values in encrypted format. Data querying and retrieval operations are carried out in encrypted format. The data values are maintained under the cloud data center or data server environment. Order Preserving Encryption (OPE) schemes are used to protect the keywords in the cloud data center. One to Many Order Preserving Encryption scheme is applied to protect the distribution attacks in the encrypted data environment. The One to Many OPE is also called ass Probabilistic OPE scheme. All the communication operations are also secured with the encryption scheme. Inverter index schemes are used to arrange the documents. The Probabilistic OPE based scheme is enhanced with security measures to handle differential attacks[12]. Term subset reassignment mechanism is integrated with the One to many OPE scheme to control change point based activities. Inverted index is protected with noise document entries to secure relevance score values. Document search and indexing operations are improved with semantic analysis methods.

The Probabilistic OPE based scheme is enhanced with security measures to handle differential attacks. Term subset reassignment mechanism is integrated with the One to many OPE scheme to control change point based activities. Inverted index is protected with noise document entries to secure relevance score values. Document search and indexing operations are improved with semantic analysis methods. Attack analysis and protection operations are integrated with the Probabilistic OPE technique. Conceptual relationship based search scheme is adapted in the encrypted data search process. Query process is carried out with privacy preserved manner. The system is divided into six major modules. They are Cloud Server, Relevance Score Assignment, Probabilistic OPE, Attack Analysis, Index Distribution Security and Query Process.

Cloud server manages the encrypted data values. Relevance score assignment module is designed to update weight values. Data encryption is carried out under the Probabilistic OPE module. Attack analysis module is used to discover the differential attacks. Index values are protected using index distribution security process. Query process is called to perform encrypted data search process.

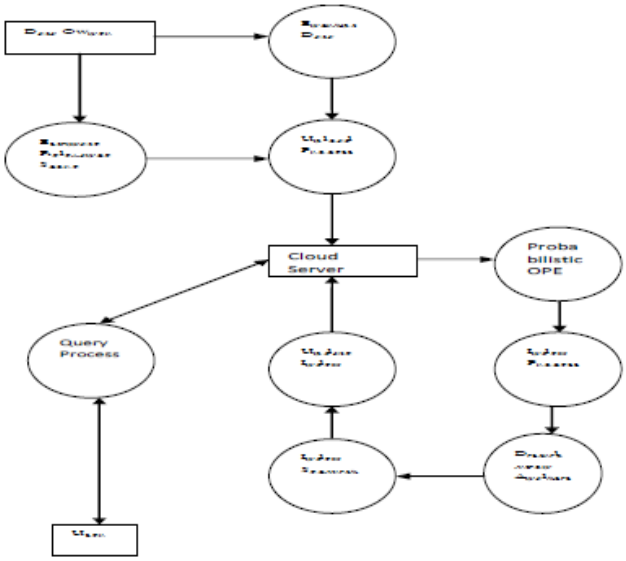


Fig: 5.1. High Performance Data Search with

Probabilistic Order Preserving Encryption (POPE) is employed to encrypt the relevance scores with index values. Inverter index is used to arrange the relevance score values. Weight values are also integrated with the index process. Random values are used to reassign the distribution intervals. Attack analysis is initiated to verify the index distribution levels. Differential attacks are discovered with distribution relationship values. Index subsets are analyzed in the attack analysis process. Differential attacks are discovered with query keyword intervals.

Encrypted data values are maintained under the cloud server application. Data encryption and upload operations are initiated by the data owner. Data owner and user details are managed under the cloud server. Data owner provides the key value for the users. The relevance score is assigned for the plain text values. The system integrates the relevance score with weight values. Term weights are estimated using statistical analysis. Concept relationship analysis mechanism is applied to estimate the semantic weights.

Index distribution security process is used to control differential attacks. Noise document entries are inserted to protect the relevance score and index values. Change point activities are controlled with term subset reassignment technique. The index distribution security is also applied to protect the semantic index values. The query process is initiated to search on encrypted data values. User privacy is ensured with query keyword encryption process. Query results are ranked with relevance score and weight values. Semantic relationship based search scheme is integrated

https://dx.doi.org/10.22161/ijaers/nctet.2017.11 ISSN: 2349-6495(P) | 2456-1908(O)

with the query process.

VI. PERFORMANCE ANALYSIS

The encrypted data search under cloud environment is designed to handle the security attacks against the searching process. Documents are maintained under the cloud environment in encrypted format. The keywords are managed in encrypted index structure. Relevance score values are adapted to fet ch the documents with ranked manner. The Order Preser ving Encryption (OPE) scheme is employed to handle th e data encryption and decryption process. One to Many Order Preserving Encryption scheme is designed to protect the data security in searching process. The One to Many Order Preserving Encryption is also referred as Probabilistic Order Preserving Encryption (POPE) scheme. Differential attacks are raised against the POPE scheme. The Enhanced Probabilistic Order Preserving Encry ption (EPOPE) scheme is designed to protect the data se arch process with differential attack handling mecha nism. Semantic relationship based data search scheme is also integrated with the EPOPE scheme. Statistical relati onship based data search and semantic or concept relatio nship based data search schemes are supported by the syste m.

The cloud data management system is tested with two security techniques. They are Pr obabilistic Order Preserving Encryption (POPE) and Enhan ced Probabilistic Order Preserving Encryption (EPOPE) schemes. The cloud data sharing system is designed as two applications such as cloud server and the client. T he Cloud Server manages the documents in encrypted f orm. The client application is served to handle the data se arch process. The system is tested with two performance measures query execution time and accuracy levels. The query execution time is measured with the query submitted time and data received time intervals. Figure 6.1. shows the query execution time analysis between the Probabilistic Order Preserving Encryption (POPE) sch eme and the (EPOPE)scheme. The analysis results show that the Enhanced Probabilistic Order Preser ving Encryption (EPOPE) scheme reduces the query execution time 25% than the Probabilistic Order Preserving En cryption (POPE) scheme.

The accuracy level is measured with the query keyword and retrieved document inform ation. Figure 6.2. shows the accuracy level analysis between the Probabilistic Order Preserving Encryption (POPE) scheme and the (EPOPE)scheme. The analysis results show that the Enhanced Probabilistic Order Preser ving Encryption (EPOPE) scheme increases the accurac y level analysis 15% than the Probabilistic Order Preserving Encryption (POPE) scheme.

VII. CONCLUSION AND FUTURE WORK

User data security and privacy are supported by the encrypted cloud storage services. One to many Order Preserving Encryption (OPE) is appl ied to perform document search on encrypted data collec tion. Differential attack handling mechanism is integrated with the probabilistic OPE scheme. Semantic query based indexing and document retrieval scheme is adapted to improve the search levels. The system provides query privacy in search process under encrypted cloud data services. Search duration is reduced in the semantic relationship based encrypted keyword search process. Accuracy is improved with relevance score and semantic query model. The system controls the keyword inferring attacks with change point modification and noise keyword insertion mechanism. The system can be enhanced to support user groups and personalized encr yption and key management scheme. The system can be en hanced to support distributed storage server mechanism for document storages.

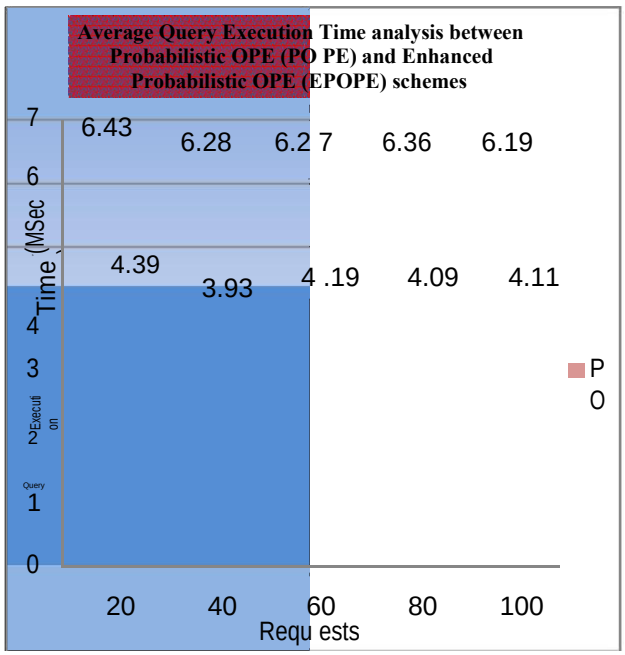
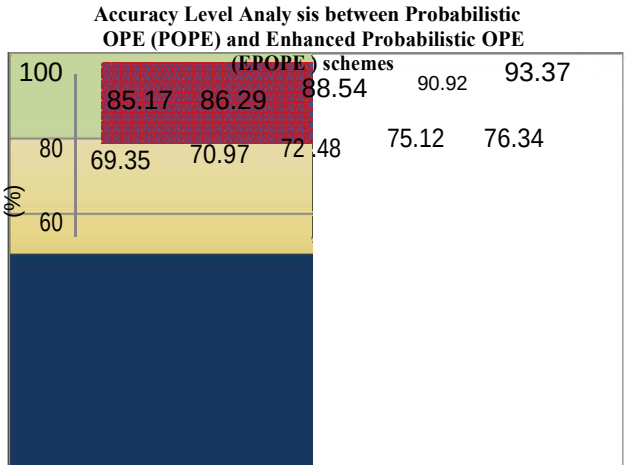


Fig:6.1. Average Q uery Execution Time analysis between Probabilistic OP E (POPE) and Enhanced Probabilistic OPE (EPOPE) schemes



https://dx.doi.org/10.22161/ijaers/nctet.2017.11 *ISSN: 2349-6495(P) | 2456-1908(O)*

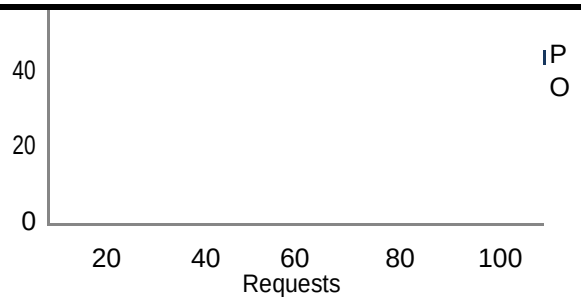


Fig: 6.2. Accura cy Level Analysis between Probabilistic OPE (POPE) and Enhanced Probabilistic OPE (EPOP E) schemes

REFERENCES

[1]. P.A. Cabarcos, F.A. Mendoza, R.S. Guerrero, A.M. Lopez, and D. Diaz- S anchez, "SuSSo: seamless and ubiquitous single sign-on for cloud service continuity across device s," IEEE Trans. Consumer Electron.,vol. 58, no. 4,pp. 1425-1433, 2012.

[2]. D. Diaz-Sanchez, F. Almenarez, A. Marin, D. Proserpio, and P.A. Cabarcos "Media cloud: an open cloud computing middle ware for content management," IEEE Trans. Con sumer Electron., vol. 57, no. 2, pp. 970-978, 2011.

[3]. S. G. Lee, D. Lee, and S. Lee, "P ersonalized DTV program recommendation syste m under a cloud computing environment," IEEE Trans. Consumer Electron., vol. 56, no. 2, pp. 1034 -1042, 2010.

[4]. W. Sun, B. Wang, N. Cao, M. Li , W. Lou, and Y. T. Hou, "Privacy preserving m ulti-keyword text search in the cloud supporting similarity based ranking," ASIACCS 2013, Hangzhou, China, May 2013, pp. 71-82, 2013.

[5]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multikey wo rd ranked search over encrypted cloud data," Pro ceedings of IEEE INFOCOM 2011, pp. 829-837, 2 011.

[6]. Q. Chai, and G. Gong, "Veri fiable symmetric searchable encryption for semi-honest-but-curious cloud servers," Proceedings of I EEE International Conference on Communications (ICC'12), pp. 917-922, 2012.

[7]. S. Grzonkowski, and P. M. Co rcoran, "Sharing cloud services: user authentication for social enhancement of home networki ng," IEEE Trans. Consumer Electron., vol. 57, n o. 3, pp. 1424-1432, 2011.

[8]. R. Sanchez, F. Almenares, P. Arias, D. Diaz-Sanchez, and A. Marin, "Enhancing privacy and dynamic federation in IdM for consumer cloud computing," IEEE Trans. Consumer Electron., vol. 58, no. 1, pp. 95-103, 2012.

[9]. J. Li, Q. Wang, C. Wang, N. Cao , K. Ren, and W. Lou, "Fuzzy keyword search ov er encrypted data in cloud computing," Proceedings of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, pp. 1-5, Mar. 2010.

[10]. C. Wang, N. Cao, J. Li, K. R en, and W. Lou, "Secure ranked keyword search over encrypted cloud data," Proceedings of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), pp . 253-262, 2010.

[11]. V Janardhan Babu, B. Suresh Babu, P. Jayanti "Optimal Redundancy Design of Stochastic LAN Devices Failures in Private Cl ouds" published paper in International Journal of Scientific & Engineering Research, Volume 8, issue 5, pp 95-99, 2017

[12]. Jeelan S, Janardhan Babu V, Vijai S, " Theft Prevention System: An appro ach to prevent Theft" Journal of Computing Te chnologies, ISSN :2278-3814, Vol 1, Issue : 2 June 2012