Research Inventy: International Journal of Engineering And Science
Vol.4, Issue 6 (June 2014), PP -46-52
Issn (e): 2278-4721, Issn (p):2319-6483, www.researchinventy.com

Privacy Preserving Ranked Multi-Keyword Search for Multiple Data Owners in Cloud Computing

J.Satya Kiran[1], Vmr Kiran[2], Mrs. Venkata Ram Reddy[3]
[1,2,3] Computer Science Engineering Department, Sree Dattha Institute of Engineering & Science

Abstract: *With the advent of cloud computing, it has become increasingly popular for data owners to outsource their data to public cloud servers while allowing data users to retrieve this data. For privacy concerns, secure searches over encrypted cloud data has motivated several research works under the single owner model. However, most cloud servers in practice do not just serve one owner; instead, they support multiple owners to share the benefits brought by cloud computing. In this paper, we propose schemes to deal with Privacy preserving Ranked Multi-keyword Search in a Multi-owner model (PRMSM). To enable cloud servers to perform secure search without knowing the actual data of both keywords and trapdoors, we systematically construct a novel secure search protocol. To rank the search results and preserve the privacy of relevance scores between keywords and files, we propose a novel Additive Order and Privacy Preserving Function family. To prevent the attackers from eavesdropping secret keys and pretending to be legal data users submitting searches, we propose a novel dynamic secret key generation protocol and a new data user authentication protocol. Furthermore, PRMSM supports efficient data user revocation. Extensive experiments on real-world datasets confirm the efficacy and efficiency of PRMSM.*

Index Terms*: Cloud computing, ranked keyword search, multiple owners, privacy preserving, dynamic secret key*

I. Introduction

The CSP itself, since the CSP possesses full control of cloud hardware, software, and owners' data. Encrypt Cloud computing is a subversive technology that on sensitive data before outsourcing can preserve is changing the way IT hardware and software are. data privacy against CSP. However, data encryption designed and purchased [1]. As a new model of makes the traditional data utilization service based on computing, cloud computing provides abundant benefits including easy access, decreased costs, quick. A trivial solution to this problem is to download all deployment and flexible resource management, etc. the encrypted data and decrypt them locally. However, this method is obviously impractical because it will cause a huge amount of communication overhead. Therefore, developing a secure search service over encrypted cloud data is of paramount importance .Secure search over encrypted data has recently at- including emails, personal health records and govern-tracked the interest of many researchers. Song et al. ment confidential files, to the cloud. This is because [3] first define and solve the problem of secure search once sensitive data are outsourced to a remote cloud, over encrypted data. They propose the conception of the corresponding data owners lose direct control of searchable encryption, which is a cryptographic prim-these data [2]. Cloud service providers (CSPs) that enables users to perform a keyword-based promise to ensure owners' data security using mecha-search on an encrypted dataset, just as on a plaintext mechanisms like virtualization and firewalls. However, these dataset expected under the circumstances" security surety contrasted with past searchable symmetric encryption (SSE) plans. 3.Extensive exploratory results exhibit the adequacy and proficiency of the proposed arrangement.

II. Problem Formulation

We present a formal description for the target problem in this paper. We first define a system model and a corresponding threat model. Then we elucidate the design goals of our solution scheme and a list of notations used in later discussions. We first define a system between keywords and files, we propose a new additive model and a corresponding threat model. Then goals of our solution scheme and which helps the cloud server return the most relevant a list of notations used in later discussions. search results to data users without revealing any sensitive information. To prevent the attackers from eavesdropping secret keys and pretending to be legal

2.1 System Model

data users submitting searches, we propose a novel In our multi-owner and multi-user cloud computing model, four entities are involved, as illustrated - dynamic secret key generation protocol and a new data user authentication protocol. As result, attackers in Fig. 1; they are data owners, the cloud server, 0018-9340 (c)

III. Data User Authentication

randomly chooses a keyword w∗ , encrypts it to ˆ w∗ , and sends ˆ w∗ to A. A outputs its guess w′ for w∗ , To prevent attackers from pretending to be legal data and wins if w′ = w∗ . We define the probability users performing searches and launching statistical that A breaks keyword secrecy as Adv attacks based on the search result, data users must A=Pr[w′ = w∗]. We say that PRMSM achieves keyword secrecy if be authenticated before the administration server re- Adv encrypts trapdoors for data users. Traditional authentication methods often follow three steps. First, data and u denotes the size of keyword dictionary A= 1+ ε, where ϵ is a negligible parameter, t u−t denotes the number of keywords that A has known, requester and data authenticator share a secret key, say, k 0. Second, the requester encrypts his personally identifiable information d 0 using k 0 and sends the en- 2.4 Notations crypted data (d 0) to the authenticator. Third, the au- thenticator decrypts the received data with k 0 and authenticates the decrypted data. However, this method k 0 O: the data owner collection, denoted as a set of • Fi: the plaintext file collection of Oi, denoted as has two main drawbacks. First, since the secret key a set of n data file Fi=(Fi, 1, Fi, 2,. . . , Fi,n). shared between the requester and the authenticator. Ci: the ciphertext file collection of Fi, denoted as remains unchanged, it is easy to incur replay attack. m data owners O = (O 1 , O 2 , . . . , Om). Ci=(Ci, 1, Ci, 2,. . . , Ci,n).

Second, once the secret key is revealed to attackers,

• W: the keyword collection, denoted as a set of u the authenticator cannot distinguish between the legal keywords W = (w 1 , w 2 , . . . , wu). requester and the attackers; the attackers can pretend

• Wi: Oi’s encrypted keyword collection of W, to be legal requesters without being detected.

Preliminaries

Now we give an example to illustrate the main idea of Before we introduce our detailed construction, we first the user authentication protocol(the detailed protocol briefly introduce some techniques that will be used in is elaborated in the following subsections). Assume this paper. Alice wants to be authenticated by the administration server, so she starts a conversation with the server. The server then authenticates the contents of the conversation.

Bilinear Map

If the contents are authenticated, both Let G and G 1 denote two cyclic groups with a prime Alice and the server will generate the initial secret key order p. We further denote g and g 1 as the generator according to the conversation contents. After the initialization, to be authenticated successfully, Alice has G and G 1, respectively. Let ˆ e be a bilinear map e : G×G → G 1, then the following three conditions are to provde the historical data of their conversations. satisfied:

1) Bilinear: ∀a, b ∈ Z ∗ p, ˆ e(ga, gb) = ˆ e(g, g) ab. 2) If the authentication is successful, both Alice and the Non-degenerate: ˆ e(g, g)≠ 1. 3) Computable: ˆ e can be administration server will change their secret keys efficiently computed. In this way, the secret keys keep changing dynamically; without the contents of the conversation

3.2 Bilinear Diffie-Hellman Problem and Bilinear

knowing the correct historical data, an attacker cannot

Diffie-Hellman Assumption start a successful conversation with the administration server. The Bilinear Diffie-Hellman (BDH) problem in (G, G 1 , ˆ e) is described as follows, given random g ∈ G, and ga, gb, gc for some a, b, c ∈ Z ∗

4.2User Authentication p, compute ˆ e(g, g) abc ∈ G 1. Before we introduce the dynamic key generation The BDH assumption is presented as follows, given method and the authentication protocol, we first intro- (G, G 1 , ˆ e), g ∈ G, and ga, gb, gc for some a, b, c ∈ Z ∗ p, duce the format of the authentication data. As shown

IV. Literature Survey

The encryption on information is a powerful approach to secure the classification of information in cloud. Be that as it may, with regards to looking, effectiveness gets low. In writing numerous exploration works are not effective in scanning uncommonly for complex inquiries. This wastefulness may prompt spillage of significant data to unapproved people groups. Tune et al, surprisingly proposed the down to earth symmetric searchable system taking into account cryptography. In this plan the record is encoded word by word. To look for a watchword client sends the catchphrase with same key to the cloud. The downside of this plan is that the word recurrence will be uncovered. Goh et al attempted to defeat the downside of Song's plan by developing secure file table utilizing pseudorandom capacities and one of a kind archive identifier randomized sprout channels. Bosch et al dealt with the idea given by Goh et al. what's more, presented the idea of special case seeks. The disadvantage of this plan is that blossom channels may present false positives. In Chang's et al proposed plan, a record is manufactured for every report. The plan is more secured contrasted with Goh's plan since number of words in a document is not unveiled. The constraint of this plan is that it is less effective and does not bolster self-assertive upgrades with new words. Golle et al plan permits various watchword looks with one scrambled question. Be that as it may, this plan is not functional. Curtmola et al surprisingly proposed the idea of symmetric searchable encryption (SSE), later on Kamara et al proposed an augmented form of SSE called dynamic SSE (DSSE), where expansion and erasure of records can be performed in file table. Every one of these plans depend on single watchword hunt.

The principal open key encryption with watchword look (PEKS) was proposed by Boneh et al. The plan experiences derivation assault on trapdoor encryption technique. Baek et al, Rhee et al enhanced hardness of security of Boneh's plan. Baek's plan presents the idea of conjunction of catchphrase inquiry. General society key encryption routines are computationally tedious and complex that makes these calculations wasteful. In Yang et al conspire the scrambled information is sought by individual clients utilizing an one of a kind key designated to them. The plan experiences key administration. Boneh et al talked about useful encryption and identified with conjunctional pursuit, extent inquiries and subset questions. Katz et al plan is a redesigned variant of Boneh's plan and examined predicate encryption for inward items and backings both conjunctions and disjunctions look on scrambled information.

There are numerous looking procedures executed in the cloud. These strategies bolster just correct watchword look. Utilizing fluffy hunt the precise catchphrases are shown alongside likeness watchwords and is dissected in [8]. This work focuses on taking care of the issues of the client who seeks the information with the assistance of fluffy watchword on cloud. Here we proposed a system where a modified file (executed utilizing connected rundown) having archive identifiers is kept up for every watchword. Each hub in the rundown stores data about the position and the decoding key of the following hub. The hubs from every single modified record are encoded with irregular keys and are haphazardly embedded into a cluster. With this, by knowing position and unscrambling key of the first hub of a rearranged list, it is conceivable to discover all archives which incorporate the relating watchword. To enhance the productivity of the above plan, top-k single watchword recovery plans are proposed in the writing .

Much work has been done in security saving multi-catchphrase hunt on scrambled information down distributed computing segment. In [11], a model is recommended that takes care of the issue of successful secure positioned catchphrase seek over encoded cloud information. Here, it proposes a current cryptographic primitive, request safeguarding symmetric encryption (OPSE). The impediments of this system are: does not bolster multi-catchphrase, does exclude IDF (characterize) for the figuring of scores, does not utilize progressed crypto strategies. Present the first strategy that gives positioned results from multi-watchword looks on open key scrambled information. By maintaining a strategic distance from a straight output of the records and by parallelizing the calculations to the conceivable degree, this system decreases the computational many-sided quality of open key cryptosystem. The plan encodes catchphrase data of every archive in a blossom channel, and progressively total (utilizing homomorphic encryption) the individual records into a tree structure. Customer will do the question handling, and cross the tree in best-first way. The question is escaped the server or cloud supplier by utilizing a productive private data recovery (PIR) convention. In this system the records are split into numerous pieces, and utilize a few CPUs in parallel to execute the client inquiries effectively. MRSE plan that deals with likeness based positioning. Here pursuit list is made on the premise of term recurrence and vector space. Quest file is utilized for multi watchword hunt and positioning the query item. Look effectiveness is enhanced by applying tree structure on record.

The future work being multi-watchword semantic hunt over the scrambled information has been spoken to in [6]. Considering the vast number of information clients and reports in the cloud, it is important to permit numerous watchwords in the inquiry demand and return records in the request of their significance to these catchphrases. Here, security protecting multi-catchphrase positioned look over scrambled information in distributed computing (MRSE) is proposed where among different multi-watchword semantics, it picks the proficient similitude measure of direction coordinating and subsequently utilizes the cryptographic procedures. Thusly, it needs respectability check of rank request in query item and protection in more grounded danger model. Equivalent word based numerous catchphrases positioned seek over scrambled cloud information utilizing adjusted paired tree is proposed in [15]. Here creator utilized symmetric encryption strategy for outlining searchable encryption plot and utilized b-tree for indexing.

Albeit numerous analysts over the globe have been exploring to distinguish a suitable security safeguarding strategy for cloud space, none of these arrangements ensure 100 percent protection. There exists an extensive variety of exploration difficulties. We thusly worked towards meeting this test.

V. Problem Formulation

Searchable Encryption (SE) plans keep up the classification and security of proprietor's information by encouraging looking watchwords specifically on scrambled information. Clients can transfer their scrambled information to cloud. Later, the approved clients can perform private watchword look on scrambled information in cloud. Different spaces like cryptography, indexing, stockpiling and so on are included in formulating proficient, secure, SE calculations over encoded documents. The members of a safe inquiry model in a cloud, normally includes information proprietor, information client and cloud server. Information proprietor scrambles the documents and using so as to relate catchphrases based file records any known cryptographic calculations. Both the encoded records and list documents are transferred to the cloud server. The trapdoors (encoded catchphrases) are utilized to inquiry scrambled records by cloud server in cloud database.

A. Framework Model

Our framework comprises of 3 elements information proprietor, information client and the cloud server as appeared in Figure 1.

1 .Data proprietor scrambles the information records for securing the information in cloud utilizing Commutative RSA (CRS) before transferring into the cloud. They likewise characterize the entrance rights for the client who need to get to those archives. The entrance right is a 2-state variable: consent conceded or authorization denied. Information proprietor makes a record tree in view of B tree and encodes the tree utilizing CRS.

2. Cloud server stores the scrambled information records and encoded file tree. It acknowledges the scrambled watchwords (trapdoor) and gives back the coordinating information record in view of their pertinence.

Information client can look for encoded information documents in cloud with scrambled catchphrases (trapdoor). The reason for utilizing encoded catchphrases is that even the cloud server must not have the capacity to induce the substance of information documents.

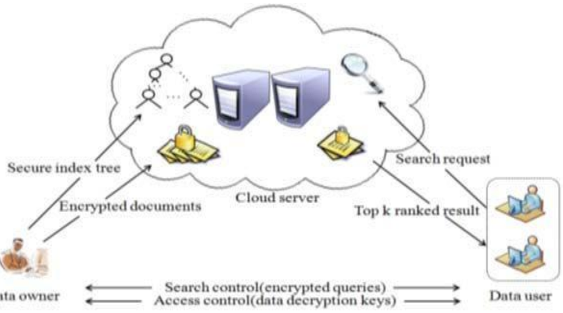

Figure 1: Searchable Encryption Architecture using CRS

B. Risk Model

The risk model for our hunt plan embraces "legitimate however inquisitive" cloud server, that is the cloud server "sincerely" takes after the convention determination, yet it is "interested" to gather and break down information (counting files) in its stockpiling and message streams got amid the convention to learn extra data.

C. Plan Goals

The proposed arrangement addresses the accompanying necessities

1. The pursuit on scrambled archive/record must be completely secure and cloud server must not have the capacity to induce the substance of the reports in any capacity.

2. The indexed lists must be positioned all together of importance .To empower positioned searchable encryption for viable usage of outsourced and scrambled cloud information under the previously stated model, our framework configuration ought to accomplish the accompanying security and execution ensure. In particular, we have the accompanying objectives: 1) Ranked catchphrase hunt: to investigate diverse instruments for planning successful positioned pursuit plans taking into account the current searchable encryption system; 2) Security certification: to keep cloud server from taking in the plaintext of either the information records or the sought watchwords, and accomplish the "as-solid as could be expected under the circumstances" security quality contrasted with existing searchable encryption plans; 3) Efficiency: above objectives ought to be accomplished with least correspondence and calculation overhead.

Evaluation Settings
That, given the same number of files (n=1000), SRMSM We conduct performance experiments on a real data and PRMSM consume much less time than MRSE set, the Internet Request For Comments dataset (RFC) on constructing indexes. Additionally, SRMSM and [28]. We use Hermetic Word Frequency Counter [29] PRMSM are insensitive to the size of the keyword to extract keywords from each RFC file. After the dictionary for index construction, while MRSE suf- keyword extraction, we compute keyword statistics fers a quadratic growth with the size of keyword such as the keyword frequency in each file, the length dictionary increases. Fig. 6(c) shows the encoding of each file, the number of files containing a specific efficiency of our proposed AOPPF. The time spent on keyword, etc. We further calculate the relevance score encoding increases from 0.1s to 1s when the number of a keyword to a file based on these statistics. The of keywords increases from 1000 to 10000. This time file size and keyword frequency of this data set can cost can be acceptable. be seen in [20]. 8.2.2 Trapdoor Generation Theexperiment programs are coded using the Python programming language on a PC with 2.2GHZ Compared with index construction, trapdoor genera- Intel Core CPU and 2GB memory. We implement tion consumes relatively less time. Fig. 7(a) demon- all necessary routines for data owners to preprocess strates that, given the same number of queried key- data files: for the data user to generate trapdoors, words (q=100), SRMSM and PRMSM are insensitive to for the administrative server to re-encrypt keywords, the size of keyword dictionary on trapdoor generation trapdoors, and for the cloud server to perform ranked and consumes 0.026s and 0.031s, respectively. Mean- searches. We use the Weil pairing [30] to construct our while, MRSE increases from 0.04s to 6.2s. Fig. 7(b) bilinear map.

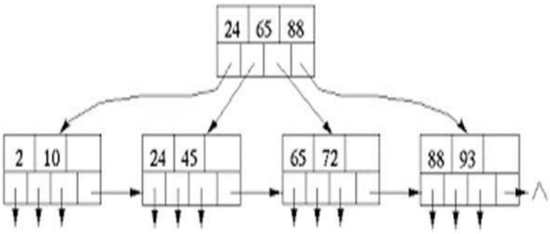

Figure 2: B tree data structure

To outline an effective multi-watchword searchable encryption plan in light of open key cryptography, we incorporated the accompanying modules. Encryption Module: By utilizing CRS, information in a record can be redesigned progressively without influencing the general execution of seeking on B-tree. On the off chance that the scrambled filed information is changed, re-indexing for the entire information is not required. Correspondingly there is no need of re-encoding the records in the database at whatever point the document is adjusted. This is an attractive element as it lessens the calculation time.
Information proprietor first produces mystery and open key pair (EK, DK) utilizing a standard open key encryption plot ie CRS. At that point proprietor makes people in general key DK open and keeps the mystery keys EK private. Archives {D | D1, D2,… , Dn} are encoded utilizing EK coming about as a part of a ciphertexts {C | C1,C2,… .Cn}. The created C is put away in cloud database.

The built file taking into account B tree is additionally scrambled utilizing CRS, i.e each inferred watchwords {W| w1,w2,… .wn}from a record is ordered in a tree **9**

VI. Related Work

Fig. 8(a) illustrates the re-encryption time cost of the In this section, we review three categories of work: administration server in PRMSM. As we can see, for searchable encryption, secure keyword search in cloud the same average number of keywords per owner, the computing, and order preserving encryption. more data owners are involved, the more time is spent on re-encryption. When there are 300 data owners,

Searchable Encryption

Search data owner has 100 keywords; we need 3.8s to The earliest attempt of searchable encryption was re-encrypt these keywords, which is acceptable. made by Song et al. In [3], they propose to encrypt Fig. 8(b) demonstrates the time cost of re-encrypting each word in a file independently and allow the server trapdoors. We observe that, for the same average to find whether a single queried keyword is contained number of trapdoors per user, the more data users in the file without knowing the exact word. This submit trapdoors, the more time would be spent proposal is more of theoretic interests because of high on re-encryption. When there are 1000 data users computational costs. Goh et al. propose building a who concurrently submit data, each data user has 10 keyword index for each file and using Bloom filter trapdoors; we only need 3.34s for re-encryption. to accelerate the search [4]. Curtmola et al. propose building indices for each keyword, and use hash tables as an alternative approach to searchable en- 8.2.4 Search cryption [5]. The first public key scheme for keyword From Fig. 9, we observe that, PRMSM spends more search over encrypted data is presented in [6]. [7] and time for searching. The fundamental reason is that, [8] further enrich the search functionalities of search

The pairing operation used in PRMSM needs more able encryption by proposing schemes for conjunctive time. As we can see from Fig. 9(a) and Fig. 9(c), keyword search. the more keywords existing in the cloud server, the The searchable encryption cares mostly about single more time is required for pairing operation. Fig. 9(b) keyword search or boolean keyword search. Extend- confirms that when the number of keywords stored on ing these techniques for ranked multi-keyword search the cloud server remains a constant, PRMSM will not will incur heavy computation and storage costs. increase even if the number of files increases. Though PRMSM spends relatively more time, this observation

Secure Keyword Search in Cloud Computing

Also confirms that the searching operation should be The privacy concerns in cloud computing motivate outsourced to the cloud server. the study on secure keyword search. Wang et al.

VII. Performance Analysis

The security of the outlined framework is given by utilizing CRS. For whatever length of time that private key (scrambled) is kept mystery the cloud supplier can't conclude list tree or records set. Since trapdoor is additionally scrambled utilizing CRS, the supplier can't make out the catchphrases inside the trapdoor keeping up the classification at list and inquiry level. The archives in distributed storage are additionally ensured, since records are scrambled utilizing CRS. Without having the decoding key it is exceptionally difficult to unscramble the reports in this manner gives security at capacity level. To be helpful and usable, databases must bolster operations, for example, inquiry, cancellation and insertion of information. For extensive associations the databases are gigantic in size and can't be kept up altogether in memory.

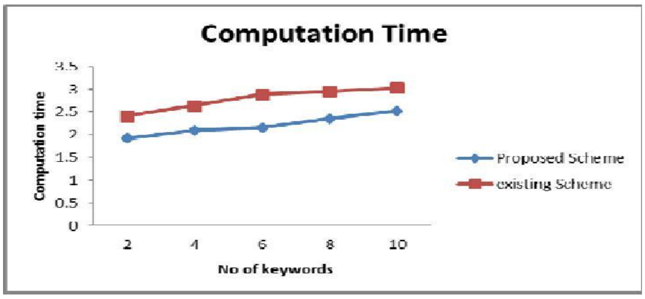

Figure 3: Time Comparison

VIII. Conclusions

Avoids the ranking order being distorted by several high frequency keywords [13]. Li et al. In this paper, we explore the problem of secure [14], Chuah et al. [15], Xu et al. [16] and Wang et al.multi-keyword search for multiple data owners and [17] proposed fuzzy keyword search over encrypted multiple data users in the cloud computing environment the contents of the conversation -cloud data aiming at tolerance of both minor type Different from prior works, our schemes and format inconsistencies for users' search input. [19] enable authenticated data users to achieve secure, further proposed privacy-assured similarity search convenient, and efficient searches over multiple data. Mechanisms over outsourced cloud data. In [20], we owners' data. To efficiently authenticate data users proposed a secure, efficient, and distributed keyword and detect attackers who steal the secret key and search protocol in the geo-distributed cloud environ-perform illegal searches, we propose a novel dynamic. secret key generation protocol and a new data user The system model of these previous works only authentication protocol. To enable the cloud server to consider one data owner, which implies that in their perform secure search among multiple owners' data solutions, the data owner and data users can easily encrypted with different secret keys, we systematical-communicate and exchange secret information. When construct a novel secure search protocol. To rank the numerous data owners are involved in the system, search results and preserve the privacy of relevance secret information exchanging will cause considerable scores between keywords and files, we propose a communication overhead. Sun et al. [21] and Zheng novel Additive Order and Privacy Preserving Function- et al. [22] proposed secure attribute-based keyword family. Moreover, we show that our approach search schemes in the challenging scenario where is computationally efficient, even for large data and multiple owners are involved. However, applying CP-keyword sets. As our future work, on one hand, we ABE in the cloud system would introduce problems will consider the problem of secure fuzzy keyword for data user revocation, i.e., the cloud has to update search in a multi-owner paradigm. On the other hand, the large amount of data stored on it for a data user we plan to implement our scheme on the commercialrevocation [32]. Additionally, they do not support clouds. privacy preserving ranked multi-keyword search. Ourpaper differs from previous studies regarding the –

Acknowledgments

Paper differs from previous studies regarding the emphasis of multiple data owners in the system model. This paper seeks a solution scheme to maximally relax This work is supported in part by the National Natural- the requirements for data owners and users, so that Science Foundation of China (Project No. 61173038, the scheme could be suitable for a large number of 61472125, 61300217), and the China Scholarship Council-cloud computing users. A preliminary version of this paper appeared in the Proceedings of the IEEE/IFIP International Conference on Dependable Systems and Networks

Order Preserving Encryption

(DSN'14) [27]. The order preserving encryption is used to prevent the cloud server from knowing the exact relevance.

References

[1]. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, Agrawal et al. proposed an Order Preserving sym- A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and metric Encryption (OPE) scheme where the numerical M. Zaharia, "A view of cloud computing," ommunication of order of plain texts are preserved [33]. Boldyreva et the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[2]. C. Wang, S. S. Chow, Q. Wang, K. Ren, and W. Lou, "Privacy- al. further introduced a modular order preserving preserving public auditing for secure cloud storage," Comput- encryption in [34]. Yi et al [35] proposed an order p- ers, IEEE Transactions on, vol. 62, no. 2, pp. 362–375, 2013. reserving function to encode data in sensor networks.

[3]. D.Song, D.Wagner, and A.Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE International Sym- Popa et al. [36] recently proposed an ideal-secure posium on Security and Privacy (S&P'00), Nagoya, Japan, Jan. order-preserving encryption scheme. Kerschbaum et 2000, pp. 44–55. 0018-9340 (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information [12]Wenjun Lu; Varna, A.L.; Min Wu, 'Confidentiality-Preserving Image Search: A Comparative Study Between Homomorphic Encryption and Distance-Preserving Randomization,' Access, IEEE, vol.2, no., pp.125,141,