IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 32-39
www.iosrjournals.org

AN EFFECTIVE APPROACH FOR SECURED AND SCALABLE DATA STORAGE AND DATA RETRIEVAL IN CLOUD COMPUTING

Sachu.P.Sahi[1], Silpa Kamalan[2], Dr. Suvanam Sasidhar Babu [3]
[1.] *Department of CSE, SNGCE, Kadayiruppu, Kerala, India,*
[2.] *Department of CSE, SNGCE, Kadayiruppu, Kerala, India,*
[3.] *Department of CSE, SNGCE, Kadayiruppu, Kerala, India,*

Abstract: *Cloud computing is an important interactive**1. Introduction**paradigm to store user's data remotely in an online Cloud computing is emerging due to the cloud server. Privacy, Data security and storageare top concerns for the cloud environments. Thereprovisioning of elastic, flexible, and on demandare so many security and storage concerns regardsstorage and computing services for customers. with the cloud. So here propose a MediatedOrganizations with a low budget can now utilize certificateless public key encryption along with thecloud storage services without heavily investing in data deduplication technique and query servicesinfrastructure and maintenance. However, the loss for secure data sharing and minimizing the storageof control over data issues many security concerns space. In this paper, the data owner encrypts thefor organizations. The loss of control over data and data using the cloud generated users' public keysthestorageplatformalsomotivatescloud customers to maintain the access control over data. basedonits accesscontrolpolicies. The cloud Moreover, the privacy and confidentialityof the partially decrypts the encrypted data for the users datais an importantfactorsregarding with the The users subsequently fully decrypt the partiallycustomers. The confidentiality management by a decrypteddatausingtheirprivatekeys.The confidentiality of hecontentandthe keys iscustomer ensures that the cloud does not read any information about the customer data.*

preserved with respect to the cloud, because the Theaccesscontrol,keymanagement, cloud cannot fully decrypt the information. And encryption, and decryption processes are handled also use a data deduplication technique, it is one of by the users to ensure data security. However, importantdatacompressiontechniquesfor eliminating duplicate copies of repeating data andwhen the data are to be shared among a group, the cryptographicservices needto beflexible to has been widely used in cloud storage to reduce handle different users and manage the keys in an the amount of storage space and save bandwidth. effective manner to safeguard data confidentiality.

Andalsodealswithanefficientinformation Deduplicationis alsooneofthecritical retrievalfor rankedqueryschemesto reduce informationpackingstrategiesforremoving queryingoverhead incurred on the cloud. These duplicate copies and has been broadly utilized as a techniquesareusedtoimprovesecurityand part of Cloud storage to reduce storage space. And storageutilizationandtoretrievedifferent also allow user to retrieve files of interest from an percentages of matched files by specifying queries untrusted server without leaking any information. of different ranks. For the widespread adoption of cloud

Keywords: *Mediated Certificateless Public keystorage services, the cloud storage model should cryptography, Security Mediated, Key Generationsolve the critical issue of data confidentiality. That is, shared data mustbestronglysecured from Centre.*

I. Introduction

unauthorizedaccesses.Inordertoassure confidentiality of sensitive data stored in clouds, a, into the cloud by the users and also use an, commonly adopted approach is to encrypt the data, encryption and querying scheme for protecting the, before uploading it to the cloud. Since the cloud, data stored in the cloud., , does not know the keys used to encrypt the data,, , A novel mediated Certificateless Public, the confidentiality of the data from the cloud is, Key Encryption scheme that does not utilize, assured. Based on mediated certificateless public, pairing operations is required for securely store the, key cryptosystem scheme, propose a method to, data in cloud. Since most Certificate public key, assure the confidentiality of data stored in clouds, encryption schemes are based on bilinear pairings,, while enforcing access control requirements., they, are, computationally, expensive., Clouds also offer both highly available, Scheme reduces the computational overhead by, storage and retrieval of files at relatively low costs., using a pairing free approach. Further, the, As cloud computing becomes common, an, computation costs for decryption at the users are, increasing amount of data is being stored in the, reduced as a semitrusted security mediator, loud and shared by users with specified, partially decrypts the encrypted data before the, privileges, which define the access rights of the, users decrypt. It is required to present an advanced, stored data. And also the critical challenge of, scheme to support stronger security by encrypting, cloud storage services is the management of the, the file. It also aiming at efficiently solving the, increasing volume of data and the retrieval of, problem of deduplication with differential, required data. To make data management scalable, privileges in cloud computing. The data owners, in cloud computing, Data deduplication is a, only outsource their data storage by utilizing, specialized compression technique for eliminating, public cloud while the data peration is managed, redundant copies of data in storage. The technique, in private cloud., , , is used to improve storage utilization. Instead of, , And also propose a scheme, termed, keeping multiple copies of the same content,, efficient information retrieval for ranked queries in, deduplication removes duplicated data by keeping, which each user can choose the rank of query to, only one copy and referring other redundant data, determine the percentage of matched files to be, to that copy. Data deduplication can take place at, returned on demand which protect user privacy., either the file level or the block level. For file level,

II. Literature Review,

deduplication, it removes the redundant copies of, the same file. Deduplication can also take place at, , , , , the block level, which removes duplicate blocks of, , ChengKang Chu, Sherman S. M. Chow,, data that occur in nonidentical files. Here focus, WenGuey Tzeng, Jianying Zhou [4], in this paper, only on file level deduplication., , authors explain about How to protect users data, Cloud, computing, provide, unlimited, privacy is a central question of cloud storage. With, ,virtualized resources to users as services .Today's, more mathematical tools, cryptographic schemes, cloud service suppliers supply each extremely, are getting more versatile and often involve, offered storage and massively parallel computing, multiple keys for a single application. In this, resources at comparatively low costs. Encryption, article, consider how to compress secret keys in, and decryption techniques are used in cloud for, public key cryptosystems which support, providing security to the data stored in cloud., delegation of secret keys for different ciphertext, Deduplication, is, one, among, necessary, classes in cloud storage. No matter which one,

compression techniques for eliminating duplicate, among the power set of classes, the delegate can, copies of repetition knowledge and uses, always get an aggregate key of constant size. This, differential query services for protecting user, approach is more flexible than hierarchical key, privacy., , , , , assignment which can only save spaces if all key, Data storage and data retrieval is an, holders share a similar set of privileges.,important functionality in cloud. The challenging, , Rivest, Shamir and Adleman [5], proposed, problem is that how effectively handle the data, the first publickey encryption scheme. This, storage and the data retrieval in clouds. So here, scheme was a concrete realization of a seemingly, propose to design a system for avoiding the, paradoxical conjecture of Diffie and Hellman that, redundancy in the data when uploading the data, it was, possible, for an entity (the, sender) to, securely send another entity (the receiver) a Ostrovsky which allows a user to retrieve files but message without these two entities having a pre this scheme is of high computational cost since it existing shared secret key, without any online requires the cloud to process the query using contact between them, and even without the homomorphic encryption on every file in a receiver knowing that they were about to receive a collection, allows a client to provide an untrusted message. This functionality is achieved by server with an encrypted search query. The server generating a pair of keys instead of just one: a uses the query on a stream of documents and public key that is widely distributed for returns the matching documents to the client. New encryption, and a related private key that is kept scheme for conducting private keyword search on secret and used for decryption. streaming data which requires server to client S. S. M. Chow, C. Boyd, and J. M. G. communication is been implemented and returns Nieto [6], explains a traditional public key the content of the matching documents. The cryptosystem requires a trusted Certificate previous best scheme for private stream searching Authority to issue digital certificates that bind was shown to have communication and storage users to their public keys. Because the Certificate complexity. This technique requires a small authority has to generate its own signature on each amount of metadata to be returned in addition to user's public key and manage each user's the documents. Paper also gives an alternative certificate, the overall certificate management is method for returning the necessary metadata based very expensive and complex. on a unique encrypted download. Ostrovsky Boneh and M. K. Franklin [7] tried to build Scheme: The Ostrovsky scheme is a process of IBE with key aggregation. One of their schemes accessing the files from cloud to clients. This assumes random oracles but another does not. In scheme is very query overhead as well as every their schemes, key aggregation is constrained in time accesses the broadband connection. This the sense that all key to be aggregated must come process is more costly to accessing files at every from different "identity divisions". This greatly query. Ostrovsky protocol suffers from the increases the costs of storing and transmitting problem of lack of aggregation of queries.

ciphertexts, which is impractical in many Although it ensures privacy by using Paillier

situations such as shared cloud storage. cryptosystem, it does not lower the costs incurred AlRiyami and Paterson [8] introduces the by the customers of the cloud. first certificateless public key encryption scheme M.Finiasz and K.Ramchandran [11] which presents the concept of concrete proposed two new communicationoptimal certificateless public key cryptography. They constructions. One uses ReedSolomon codes and proposed certificateless public key encryption, allows for a zeroerror, and the other is based on signature, and key exchange schemes that were irregular LDPC codes and allows for lowerconstructed using elliptic curve pairings. computation cost at the server. The above privateNing Cao, Cong Wang, Li, Ming, Kui Ren, searching schemes only support searching for OR Wenjing Lou [9] This paper presents the searching of keywords or AND of two sets of keywords.

Of Encrypted cloud data using PrivacyPreserving B. Hore, E.C. Chang, M.H. Diallo, and S. Multikeyword Ranked Search (MRSE) method. Mehrotra, [12] extended the types of queries to In this paper coordinate matching technique is support disjunctive normal forms (DNF) of used. Coordinate matching is used to find the keywords. The main drawback of existing private similarity between search query and data searching schemes is that both the computation documents. Another technique that is Inner and communication costs grow linearly with the product Similarity, also used to describe the Multi number of users executing queries. Thus, when keyword Ranked Search over Encrypted Cloud applying these schemes to a largescale cloud Data. Here four modules of searching that are environment, querying costs will be extensive.

Encrypt Module, Client Module, MultiKeyword Q. Liu, C. Tan, J. Wu, and G. Wang, [13]

Module, and Admin Module are performed over was the first to make private searching techniques encrypted cloud data. applicable to a cloud environment. However, R. Ostrovsky and W. Skeith [10], A key requires the cloud to return all of the matched files, privacy search solution was proposed by which may cause a waste of bandwidth when only a small percentage of files are of interest. COPS (Cooperative private searching).If the users queries contain common keywords; it leads to lowered cost since the queries are aggregated by the ADL. Even in the scenario of no common keywords among users queries, the merging of queries helps in considerably lowered number of round trips to the cloud, thereby lowering the overall costs. COPS protocol can send too many results leading to excessive CPU consumption on the cloud. A lot of network bandwidth is required for transferring the response buffers from the cloud.

J. Stank, A. Sorniotti, E. Androulaki, and L. Kenco [14] with postprocess deduplication, new data is first stored on the storage device and then a process at a later time will analyze the data looking for duplication. Implementations offering policy based operation can give users the ability to defer optimization on active files, or to process files based on type and location. One potential drawback is that you may unnecessarily store duplicate data for a short time which is an issue if the storage system is near full capacity.

M. Bellare, S. Keelveedhi, and T. Ristenpart [15] MessageLocked Encryption (MLE), where the key under which encryption and decryption are performed is itself derived from the message. MLE provides a way to achieve secure deduplication, a goal currently targeted by numerous cloudstorage providers. They provide definitions both for privacy and for a form of integrity that they call tag consistency. They provide ROM security analyses of a natural family of MLE schemes that includes deployed schemes. They make connections with deterministic encryption, hash functions secure on correlated inputs.

Liu, Q., Tan, C. C., Wu, J., & Wang, G. [16] in this paper authors state that the Range Query Processing with RASP encryption is convexes maintaining. A convex set is represented by range query. The processing strategies are based on multidimensional index trees, such as RTree, that handles axis aligned minimum bounding boxes An encryption technique called Order Preserving Encryption Scheme in which encrypted data on that comparison operations is directly applied, the operands are not decrypted in that process.

Wei. L & Reiter M. K [17] this Paper contribution is that the semanticsbased multi keyword ranked search technology which supports synonym queries of encrypted cloud data. The synonyms of the predefined keywords get the input from authorized cloud customers at that time search results can be achieved. Authors presented a cloud computing middleware Media Cloud for Settop boxes for classifying, searching, and delivering media inside home network and across the cloud. This approach enables the sharing of personalized content and more sophisticated networkbased services over a conventional TCP/IP infrastructure.

III. Analysis Of Problem

Data storage and data retrieval is an important functionality in cloud. The challenging problem is to how effectively handle the data storage and the data retrieval in clouds. So here propose to design a system for encryption and querying scheme for protecting the data stored in the cloud and also avoiding the redundancy in the data when uploading the data into the cloud by the users and also introduce a querying scheme.

3.1 Disadvantages of Existing System

The existing system is key aggregate cryptosystem in which files are encrypted under the identifier of cipher index and public key and this key is generated using pairing operation and only one level of decryption which ensure less security. And there is a probability that the same users upload the same content which consumes more storage spaces in cloud. And also this existing system does not provide any privacy search for the files that are uploaded, so the searching of the files is difficult in the existing system

IV. Mediated Certificateless Encryption

The mediated certificateless public key encryption scheme is a 7tuple mCLPKE=(SetUp, SetPrivateKey, SetPublicKey, SEMKeyExtract, Encrypt, SEMDecrypt, and USERDecrypt*).* The description of each algorithm is as follows.

4.1 Setup

It takes a security parameter k as input and generates system parameters params and a secret master key mk.

4.2 Set Private Key

It takes params and ID as input and returns the user’s secret value SKIDD.

4.3 Set Public Key

It takes params and a user’s secret value SKID as input and generates the user’s public key PKID.

4.4 SEMKey Extract

Each user registers its own identity and public key to the KGC; the KGC takes params, mk and user identity ID as input and outputs a SEMkey corresponding to ID required during decryption time by the SEM.

4.5 Encrypt

It takes params, a user’s identity ID, a user’s public key PKID, and a message M as inputs and returns a ciphertext CID.

4.6 SemDecrypt

It takes params, a SEMkey, and a ciphertext CID as input, and then returns a partial decrypted message C_ ID for the user.

4.7 UserDecrypt

It takes params, a user’s private key SKID, the partial decrypted message C_ ID by the SEM as input and returns a fully decrypted message M. or a special symbol $\perp$ meaning an decryption

V. Query Services

A user can retrieve different percentages of matched files by specifying queries of different ranks. The EIRQ schemes make the private searching technique more applicable to a costefficient cloud environment. However, in the EIRQ schemes, simply determine the rank of each file by the highest rank of queries it matches.

The original EIRQ scheme named as EIRQ Efficient and the extension as EIRQ Simple. The basic idea of EIQREfficient is to construct a privacypreserving mask matrix with which the cloud can filter out a certain percentage of matched files before mapping them to a buffer. The basic idea of extensions is that, for each rank i

{0,… r}, the cloud adjusts the buffer size βi and the mapping times γi to make the file survival rate qi approach 1 i/r.

5.1 The EIRQEfficient Scheme

In EIRQ efficient scheme first should determine the relationship between query rank and the percentage of matched files which it returns. If suppose ranks are classified into 0~z ranks. Rank0 queries the highest rank and rankz queries the lowest rank. Hence rank0 queries can retrieve 100% of searched file whereas rankz query can’t retrieve any file.

The EIRQ Efficient scheme should be resolved two fundamental problems. First, it determines the relationship between query rank and the percentage of matched files to be returned. Those queries are classified into 0 to r ranks. Rank0 queries have the highest rank and the Rankr queries have the lowest rank. Secondly, it determines which matched files will be returned and which will not. Here, simply fix the probability of a file being produces by the highest rank of queries matching this file. Specifically, first rank each keyword by the highest rank of queries selecting it, and then rank each file by the highest rank of its keywords. It mainly consists of four algorithms they are QueryGen, Matrix Construct, File filter and ResultDivide. The first step is the user sends the keyword and the rank of the query to the Key Generation Center by using QueryGen algorithm. Secondly Key Generation Center runs the MatrixConstruct algorithm after aggregating enough user queries, to send a mask matrix to the cloud. The mask matrix M consists that drow and rcolumn matrix, where d is the number of keywords. And the next step is, the cloud runs the FileFilter algorithm to return a buffer. The buffer contains a certain percentage of matched files to the Key Generation Center. And finally, to distribute search results to each user by the KGC runs the ResultDivide algorithm. By executing keyword searches the KGC can find out all of the files that match users‟ queries.

ALGORITHMS:
MatrixConstruct

```
For i=1 to d do
For j=1 to γ do
If  j ≤ γ l then
M  [i, j] = Epk(1)
Else

M  [i,j] = Epk(0)
```

FileFilter

```
For each file Fj stored in the cloud do For i=1 to d do
K = j mod γ; Cj= π Dic[i] £ Fj M[I,k] ; ej = cj|Fj|
Map (Cj,ej) γ times to a buffer of sizeβ
```

5.2 The EIRQSimple Scheme

The working process of EIRQ Simple is similar to EIRQ efficient scheme. The difference occur only in the MatrixConstruct and FileFilter algorithms.

ALGORITHMS: Matrix construct

```
For i=0 to γ – 1 do For j=1 to d do
If Dic[j] is in Rank i queries then
```

.

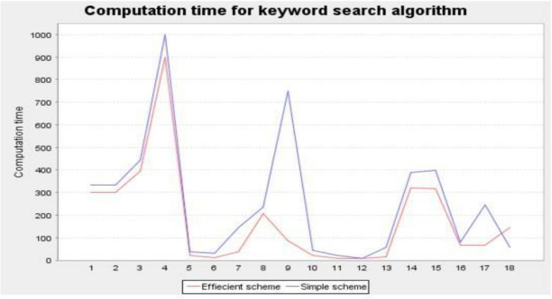

Figure.1. Keywords vs. Computation time**.**

VI. Conclusion

Qi[j]= Epk(1)
Cloud computing is used for storage,
sharing and retrieving information. Mediated

else
Certificateless Public Key Encryption is to allow a user to transmit a confidential message to another

Qi[j] =Epk (0)

user by encrypting the message using the Adjust γi and βi so that survival rate of recipient's public key which does not have to be contained in a certificate issued by Certificate Rank – i files is authority. In spite of the absence of the checking

qi =1 –i/γ

process, the sender is guaranteed that only the honest recipient who has gone through appropriate

FileFilter
authentication procedure and has obtained a right partial private key associated with his identifier ID

For i = 0 **to** γ – 1 **do**
from the Key Generation Center will be able to

For each file F in the cloud **do** decrypt the message. Data deduplication is used for avoiding the duplicated copies of the files

For j=1 to **do**

Stored in the cloud which minimizes the storage space. In EIRQ Scheme which provides

C = π Dic[j]£F Qi[j] ; e= c
|f|

Differential query services while protecting user privacy. By using this scheme user can retrieve Map (c, e) γi times to Bi of size βi different percentages of matched files by The main drawback of EIRQ simple is that specifying queries of different ranks. By further reducing the communication cost incurred on the it returns redundant files when there are files cloud, the EIRQ schemes make the private satisfying more than one ranked query. For searching technique more applicable to a cost example, if Fi is of interest by Rank0 and Rank1 efficient cloud environment. queries, it will be returned twice, which wastes the network bandwidth.

VII. Analysis

In this section, discussing about the performance analysis of proposed System. Graph shows the analysis between the EIRQ efficient scheme and simple scheme based on keywords and computational time. In analysis the EIRQ simple scheme takes more computation time compared with EIRQ efficient scheme. The Xaxis denotes queries in each rank, Yaxis denotes Computational cost.

References

[1] SeungHyun Seo, Mohamed Nabeel, Xiaoyu Ding, and Elisa Bertino,"An Efficient Certificateless Encryption for Secure Data Sharing in Public Clouds, "Ieee Transactions On Knowledge And Data Engineering, Vol. 26, No. 9, September 2014.

[2] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou,"A Hybrid Clou Approach for Secure Authorized Deduplication", Ieee Transactions On Parallel And Distributed System Vol:Pp No:99 Year 2014

[3] Qin Liu, Chiu C. Tan, Jie Wu, and Guojun Wang, "Towards Differential Query Services in CostEfficient Clouds", Ieee Transactions On Parallel And Distributed Systems, Vol. 25, No. 6, June 2014

[4] ChengKang Chu, Sherman S.M. Chow, WenGuey Tzeng, Jianying Zhou, and Robert H. Deng,"KeyAggregate Cryptosystem for Scalable Data Sharing in Cloud Storage" IEEE Transactions On Parallel And Distributed System, Vol 25, No. 2 February 2014.

[5] Rivest, R.L., Shamir, A., Adleman, L.: A method for obtaining digital signatures and publickey cryptosystems. Communications of the ACM 21, 120–126 (1978)

[6] S.S.M.Chow, C. Boyd, and J. M. G. Nieto, "Security mediated certificateless cryptography," in Proc. 9th Int. Conf. Theory Practice

PKC, New York, NY, USA, 2006, pp. 508–524.

[7] D.Boneh and M. K. Franklin, "IdentityBased Encryption from the Weil Pairing," in Proceedings of Advances in Cryptology – CRYPTO '01, ser. LNCS, vol. 2139. Springer, 2010, pp. 213–229.

[8] AlRiyami, S.S., Paterson, K.G.: CBE from CLPKE: A generic construction and efficient schemes. In: S. Vaudenay (ed.) Public Key Cryptography – PKC 2005, Lecture Notes in Computer Science, vol. 3386, pp. 398–415. SpringerVerlag (2005)

[9] Ning Cao, Cong Wang , Li, Ming , Kui Ren, Wenjing Lou, "PrivacyPreserving Multikeyword Ranked Search over Encrypted Cloud Data" INFOCOM, 2011 Proceedings IEEE April 2011.

[10] R. Ostrovsky and W. Skeith III, "Private searching on streaming data," in Proc. of ACM CRYPTO, 2005.

[11] M. Finiasz and K. Ramchandran, ''Private Stream Search at the Same Communication Cost as a Regular Search: Role of LDPC Codes,'' in Proc. IEEE ISIT, 2012, pp. 25562560.

[12] X. Yi and E. Bertino, ''Private Searching for Single and Conjunctive Keywords on Streaming Data,'' in Proc. ACM Workshop Privacy Electron. Soc., 2011, pp. 153158.

[13] Q. Liu, C. Tan, J. Wu, and G. Wang, ''Cooperative Private Searching in Clouds,'' J. Parallel Distrib. Comput., vol. 72, no. 8, pp. 10191031, Aug. 2012.

[14] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.

[15] M. Bellare, S. Keelveedhi, and T. Ristenpart. Messagelocked encryption and secure deduplication. In EUROCRYPT, pages 296–312, 2013.

[16] Liu, Q., Tan, C. C., Wu, J., & Wang, G. (2012, March). "Efficient information retrieval for ranked queries in costeffective cloud environments." In INFOCOM, 2012 Proceedings IEEE (pp. 25812585). IEEE.

[17] Wei. L & Reiter M. K. (2011). "Thirdparty DFA evaluation on encrypted files". Tech. Rep. TR11005, Department of Computer Science, University of North Carolina at Chapel Hill.