A Survey on Keyword Based Search over Outsourced Encrypted Data

S. Evangeline Sharon [#1] & N. Saravanan [*2]

[#] PG Student , School of Computing,
SASTRA University, Tirumalaisamudram, Thanjavur-613401,Tamilnadu, India.
[*] Assistant Prof., School of Computing,
SASTRA University, Tirumalaisamudram, Thanjavur-613401,Tamilnadu, India.
[1] evangelin3@gmail.com
[2] saranindia@gmail.com

***Abstract*—To ensure security, encryption techniques play a major role when data are outsourced to the cloud. Problem of retrieving the data from the cloud servers is considered. Many searching techniques are used for retrieving the data. This study focused on a set of keyword based search algorithms. It provides secure data retrieval with high efficiency. It concludes Ranked Searchable Symmetric Encryption (RSSE) scheme meant to be best methodology for searching the encrypted data.**

Keyword-Cloud, Encrypted data, Keyword based search, Outsourced data, Survey.

I.INTRODUCTION

Cloud Computing, the trendiest computing in information technology where everything is based on on-demand service and pay-for-use service. It is bringing of computing services such as SaaS, PaaS and IaaS over the internet that are supervised by arbitrator at outback locations. Many applications such as emails, file storage, business data, etc. are outsourced to cloud server. Only authorized user can access the data from the cloud server. Outsourcing unencrypted data to cloud by the owner is not much secure because server may leak information to cyberpunks. Hence encryption plays a major role before outsourcing the data into the cloud server. In spite of encrypting, retrieval of data becomes an intriguing task when searching has to be made on vast data. The best way is to use keyword based search on encrypted data for data concealing.

Many searchable techniques have been proposed on the basis of keyword search. Discussion is made on the existing techniques that are been intend by many authors. This study analyses the algorithms for searching the encrypted content. Survey is made on these algorithms based on the working principle, merits and demerits. It also compares the complexity, efficiency overhead of various algorithms and shows which technique is better to handle while retrieving the encrypted content.

II. TECHNIQUES FOR SEARCHING OVER ENCRYPTED DATA

It includes working of encryption algorithm, how searching is done on the encrypted content, advantages and disadvantages of each technique.

A. Symmetric Key Cryptography

Symmetric key cryptography works by encrypting each word in a file using two layered encryption construction. Probabilistic searching is made on the encrypted data which deals with sequential scan and indexing methodologies Provable secrecy, controllable searching, hidden queries, query isolation [1] are the four techniques which makes the algorithm efficient, simple and fast. Sequential scan meets all the above techniques but it is not effective when searching is made on huge data content. Therefore to induce searching pre-computed index plays a vital role which support advanced search queries. But to make indexing technique secure, secure index data structure [2] can be used which admits queries with a trapdoor. It is semantically secure and practicable in multi-user settings where indexes are updated frequently on the remotely located server.

B. Public Encryption Keyword Search

Public Encryption Keyword Search (PEKS),a searchable encryption technique which corresponds to symmetric key encryption .In this, file is encrypted using public key by the people who wants to store in the server but the authorized users can search a file using their private key [3]. Fig. 1 illustrates the working principle of PEKS. The four algorithms are used for this technique. First, *keyGen* is used to generate public key and private key pair for both server and user. Second, PEKS algorithm produces searchable encryption. Third, *Trapdoor* algorithm is used to calculate trapdoor with private key and keyword. Fourth, *Test* is used to match the keyword and requested word. If matches then the file are sent to the user .PEKS means Identity Based Encryption (IBE) which has major advantages such as

a) Effective system based on Diffie-Hellman problem

b) Limited system based on general trapdoor definitions

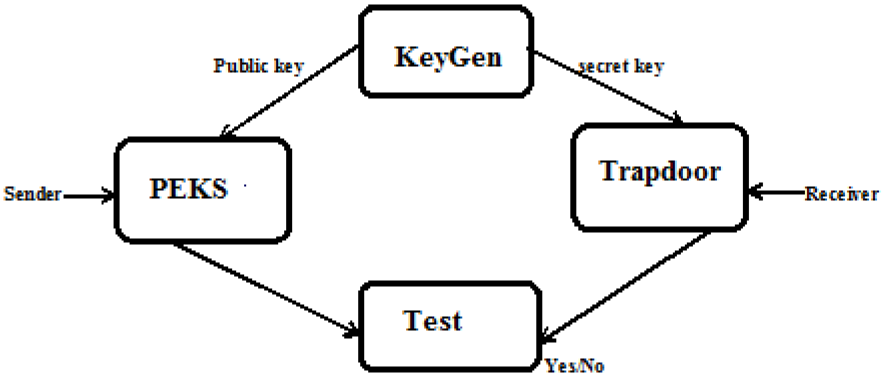

Fig 1. Working of PEKS

This scheme fails regarding access policy and dictionary attack. The major disadvantages are trapdoor contains meaningful keywords and one-one mapping takes place between trapdoor and keyword.

C. Hidden Vector Encryption

Hidden Vector Encryption (HVE) supports continuative queries [6] whereas PEKS supports only comparison and subset queries. In which it works with four algorithms namely *Setup, Encrypt, GenToken and Query* .First, *Setup* creates a bilinear group of elements using random primes and random elements, Second, *Encrypt* chooses the random element and using public key it encrypt the contents in a file. Third, *GenToken* will generate the token for the predicate using a secret key. Fourth, *Query* finds the keyword from the cipher text and if matches it return the file. Even so HVE fails for disjunctive queries because cipher text is linear to attribute.

D. Attribute Based Encryption (ABE)

Attribute based encryption (ABE) affirms one upload many download policy most formally PEKS and HVE does not support. ABE uses access policy while searching on encrypted data with its Boolean expressions. It works on the basis of nine algorithms .The first is the *Setup* algorithm used to compute secret key and master key by the trusted authority. Second, *KeyGen* algorithm is used to generate the public/private key pair. Fourth, fifth and sixth algorithm such as *PseudoGen, Encrypt, AttrScm* are used for outsourcing the data using cryptographic primitives such as access structure, bilinear maps and attribute scrambling procedure. Seventh, eighth and ninth algorithm such as query, retrieve and decrypt is mainly for the retrieval of data. In which *query* algorithm works as the retriever take on pseudonym list from the cloud service provider and receiver sends the scrambled index to the Cloud Service Provider (CSP). Then the CSP checks whether the request made by the retriever and the encrypted index stored are same by using *Retrieve* algorithm. If it matches decrypt algorithm works where the encrypted data are decrypted and sent to the retriever. It provides best quality for searching over encrypted data and faster in accessing.

E. Predicate Privacy Preserving in Public Key Encryption

Predicate privacy preserving search on keyword get the better of PEKS by using randomization technique. In which keywords are randomized and therefore trapdoors does not provide any meaningful keywords. The user and the receiver share a secret key which is not logical when there are huge number of users .To make tolerant of guessing attacks, two framework were introduced namely PEKSrand-BG for brute-force guessing and PEKSrand-SG for statistical guessing in [9].

PEKSrand-BG provides a proxy server which in advance processes the PEKS cipher text from the sender. The working principle illustrates in fig 2.

PEKSrand-SG has two methods *Proxy Farm* and *Random Walk,* in which several proxies can be maintained for storing the secret key and indirect mapping between keyword and trapdoor respectively.

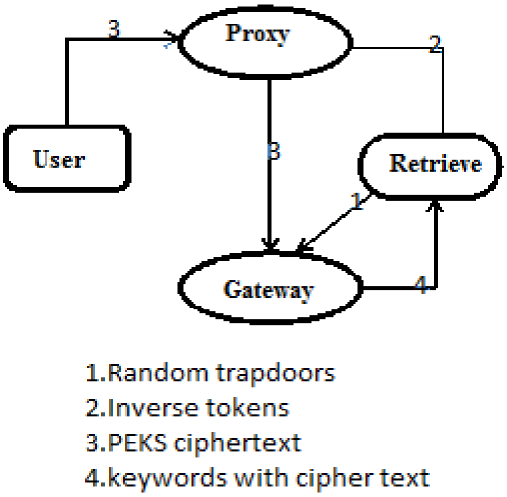

Fig 2. Working Flow of PEKSrand-BG

Therefore overall communication, computation and storage overhead are sensible when predicate privacy made in PEKS.

F. Privacy Preserving Keyword Search

It is a multi-round protocol between server and user on single keyword .It uses per index file where each document contains a keyword. The keyword index is encrypted using pseudorandom bits using heuristic pseudorandom functions [4]. On the setup phase user chooses a random secret key to encrypt the file. Then the user submits index and file content to server .On the retrieval phase, when the user wants to search or retrieve file from the server, user retrieves the index file and then computes keyword with the secret key .The computed key is sent to server, where server matches the file and then sent to the user .The per-index file scheme using pseudorandom functions is the better than using bloom filters. This scheme fails when multiple keywords are used.

G. Secure Privacy Preserving Keyword Search

Secure Privacy Preserving Keyword Search (SPKS) grants cloud service provider to decrypt the data and return file containing keywords [11]This technique overcomes the computation and communication overhead, provides query and data privacy for the users. It figures out six algorithms for efficient searching on encrypted data. The flow of SPKS is illustrated in fig 3. First, *KeyGen* used to generate a public/private key pair. Second, *EMBEnc & KWEnc* encrypts all the content in the file and keywords are encrypted respectively which then stored in the server. Third, *Tcompute* used on the retrieving phase where user generates a trapdoor and pass it to CSP. Fourth, *KWtest* checks whether the keyword contain in the encrypted data. Fifth, *PDecrypt* mainly for CSP to decrypt the intermediate result partly and sends the cipher text and the partial decrypted content .Sixth, *Recovery* runs by the user to decrypt the plain text. Therefore it provides semantic security in plain text attack.

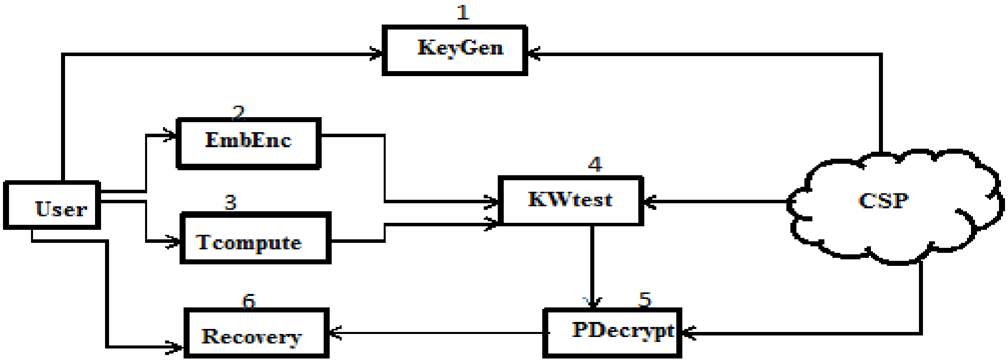

Fig 3. Working Principle of SPKS

H. Authorized Private keyword Search (APKS)

Authorized Private Keyword Search (APKS) deals with multi-keyword search, while above techniques conducts with single keyword which misses query flexibility and efficiency [10]. In fine-grained authorization framework, every user obtains searching capabilities authorization from Local Trusted Authority (LTA). Hierarchical Predicate Encryption (HPE), a cryptographic primitive uses attribute hierarchy for simple range queries.

1)APKS based on HPE

The following steps are required while searching takes place in encrypted data using multi-keyword

a. Multi-dimensional query are converted to its CNF (Conjunctive Normal Form) formula.

b. Attributes are defined in a hierarchical way. i.e., attribute hierarchy.

c. Indexes and capabilities are generated by *GenIndex* and *GenCap* algorithm respectively.

When a user wants to retrieval a file using a keyword from LTA, LTA checks whether the user has an attribute value set and if it matches then user can retrieval the file from the server. But the major disadvantage is APKS also does not prevent keyword attack.

APKS$^+$ adds a secret key while encryption and decryption takes place which hides the data from the attackers. Therefore it prevents dictionary keyword attack and accomplishes index privacy and query privacy.

I. Fuzzy Keyword Search

It enhances system usability when searching input exactly matches. Keywords are measured using edit distance and fuzzy keyword sets are constructed. Straight forward and wild card based are the two approaches are dealt with edit distance in [8].In straight forward approach edit distance are calculated where all the forms of keywords are to be listed .Based on this indexing is built .Trapdoor are shared between user and the owner .While retrieving file user computes the trapdoor based on the request, server matches with index table and return all potential identifiers. Large storage requirements and lack of efficiency are the major disadvantages. Wild card based approach overcomes the disadvantage by building a wild card fuzzy sets which calculates edit distance, keyword takes place at the same position are put together in a set.

The above all techniques based on searchable encryption supports only Boolean search which has two major drawbacks. They are,

a) User wants to decrypt every file that contains the keyword to match their file when retrieving the file is based on keyword

b) Retrieving all files leads to network traffic.

J. Keyword Search based on ranking over Encrypted Data

The major disadvantage of above mentioned techniques gets the better of in ranked keyword search. Ranked Searchable Symmetric Encryption (RSSE) framework is used to support rank search which built over the SSE cryptographic primitive. Four algorithms are used namely *KeyGen, BuildIndex, TrapdoorGen, SearchIndex*. Two phases such as setup phase which uses *KeyGen* algorithm for generating public/private key pair and *BuildIndex* algorithm to generate index file containing keywords educed from file. The file collection and the index file are outsourced after encryption with frequency based relevance score. While the retrieval phase uses *TrapdoorGen* algorithm generates a trapdoor using the user’s request. Upon the user request server runs *SearchIndex* algorithm which searches the files based on ids and relevance scores and sent files to the user. But RSSE has huge communication overhead when ranking is on user side and two round trip times is taken. Therefore efficient RSSE frame work uses Order Preserving Symmetric Encryption Scheme (OPSE) in [12]. It supports deterministic property in which a random coin generator and sampling function implemented. OPSE is used instead of encrypting scores in RSSE and in retrieval phase OPSE values are much more relevant. They provide better efficiency while retrieving files with top-k retrieval.

III. CONCLUSION

In this study rigorous analysis is made on encryption techniques which relate to search based retrieval of files from the outsourced encrypted data. Many searchable encryption schemes have been analysed based on single keyword and multi-keyword search. Many disadvantages have been focussed on these techniques since they rely on Boolean expressions. Therefore rank based retrieval of data has been discussed which proves the data security, fast search access and does not leak information to untrusted authorities. This study concludes rank based retrieval is most efficient for searching on encrypted data.

REFERENCES

[1] D.Song, D. Wagner and A. Perrig,Practical techniques for searches on encrypted data.. Proceeding of the 2000 IEEE Symposium on Security and Privacy,. May 14-17,. Washington, DC., USA.,.pp: 44-55,2000.

[2] E.J.Goh, Secure Indexes.Technical Report 2003/216.http://eprint.iacr.org/2003/216,2003.

[3] D.Boneh, G.D. Crescenzo, R. Ostrovskyand G. Persiano, Public key encryption with keyword search. Proceedings of the International Conference on the Theory and Applications of Cryptographic Techniques, May 2-6, 2004, Interlaken, Switzerland, pp: 506-522.

[4] Y.C.Chang,and M. Mitzenmacher, Privacy preserving keyword searches on remote encrypted data. Proceedings of the 3rd International Conference on Applied Cryptography and Network Security, June 7-10, 2005, New York, USA.,pp: 442-455.

[5] R.Curtmola, J.A. Garay, S. Kamara and R. Ostrovsky,Searchable symmetric encryption: Improved definitions and efficient constructions. Proceedings of the 13th ACM Conference on Computer and Communications Security, October 30-November 03, 2006, Alexandria, USA.,pp: 79-88

[6] D.Boneh and B. Waters, 2007.Conjunctive, subset and range queries on encrypted data. Proceedings of the 4th Theory of Cryptography Conference, February 21-24, 2007, Amsterdam, The Netherlands, pp: 535-554.

[7] Liu, Q., G. Wang and J. Wu, 2009. An efficient privacy preserving keyword search scheme in cloud computing. Proceedings of the International Conference on Computational Science and Engineering, Volume 2, August 29-31, 2009, Vancouver, Canada, pp: 715-720.

[8] Li, J., Q. Wang, C. Wang, N. Cao, K. Ren and W. Lou, 2010. Fuzzy keyword search over encrypted data in cloud computing. Proceedings of the 29th IEEE International Conference on Computer Communications, March 15-19, 2010, San Diego, CA., USA.,pp: 1-5.

[9] Zhu, B.and K. Ren, 2010. PEKSrand: Providing predicate privacy in public-key encryption with keyword search. Proceedings of the IEEE International Conference on Communications, June 5-9, 2011, Kyoto, Japan, pp: 1-6.

[10] Li, M., S. Yu, N. Cao and W. Lou, 2011. Authorized private keyword search over encrypted data in cloud computing. Proceedings of the 31st International Conference on Distributed Computing Systems, June 20-24, 2011, Minneapolis, MN., USA.,pp: 383-392.

[11] Liu, Q., G. Wang and J. Wu, 2012. Secure and privacy preserving keyword searching for cloud storage services. J. Network Comput. Appl., 35: 927-933.

[12] Wang, C., N. Cao, K. Ren and W. Lou, 2012.Enabling secure and efficient ranked keyword search over outsourced cloud data.Proc. IEEE Trans. Parallel Distib. Syst., 23: 1467-1479.