ISSN (Online): 2455-7838
SJIF Impact Factor (2017): 5.705

EPRA International Journal of

Research & Development

(IJRD)

Monthly Peer Reviewed & Indexed International Online Journal

Volume: 3, Issue:11,November 2018

Published By :
EPRA Journals

SJIF Impact Factor: 5.705 Volume: 3 | Issue: 11 | November | 2018 ISSN: 2455-7838(Online)
EPRA International Journal of Research and Development (IJRD)

FAST PHRASE SEARCH USING BLOOM FILTER TECHNIQUE TO RETRIEVE FILE ON CLOUD COMPUTING

M. Thulasimani [1]
[1]Research Scholar, Department of Computer Science , Vivekanandha College for Women, Tiruchengode, India,

M.Valarmathi [2]
[2]Head of the Department of Computer Science, Vivekanandha College for Women, Tiruchengode,India,

R.Limya [3]
[3]Research Scholar , Department of Computer Science , Vivekanandha College for Women, Tiruchengode, India,

ABSTRACT
Cloud computing has generated a lot of interest within the analysis community in recent years for its many benefits; however has also raise security and privacy issues. The storage and access of secret documents have been acknowledged as one of the central problems in the area. In particular, many researchers investigated solutions to look over encrypted documents keep on remote cloud servers. Here we tell about phrase ***search technique, supported Bloom filters and Nth gram technique that are quicker than existing result. In this paper now discuss about at the time of file uploading on cloud we check file deduplication. Avoid file deduplication. Therefore save the storage capacity and time. We store only unique files on cloud. At the time of file uploading we check file whether deduplicate or new file. If the file is deduplicate that will never store over the cloud. Else new file means store the documents. File reduplication checking is used for cloud storage management. Each file has unique id on the cloud. So easy to store and file retrieve.***
KEYWORDS: *Fast phrase search, Nth gram technique, Cloud, Bloom Filters.*

INTRODUCTION

Cloud computing has emerged as a disruptive trend in both IT industries and research communities recently, its salient characteristics like high scalability and pay-as you-go fashion have enabled cloud consumers to purchase the powerful computing resources as services according to their actual requirements, such that cloud users have no longer need to worry about the wasting on computing resources and the complexity on hardware platform management. Nowadays, more and more companies and individuals from a large number of big data application shave outsource their data and deploy their services into cloud servers for easy data management, efficient data mining and query processing tasks. Data encryption has been widely used for data privacy preservation in data sharing scenarios, it refers to mathematical calculation and algorithmic scheme that transform plaintext into cipher text, which is a non-readable form to Secure data deduplication can

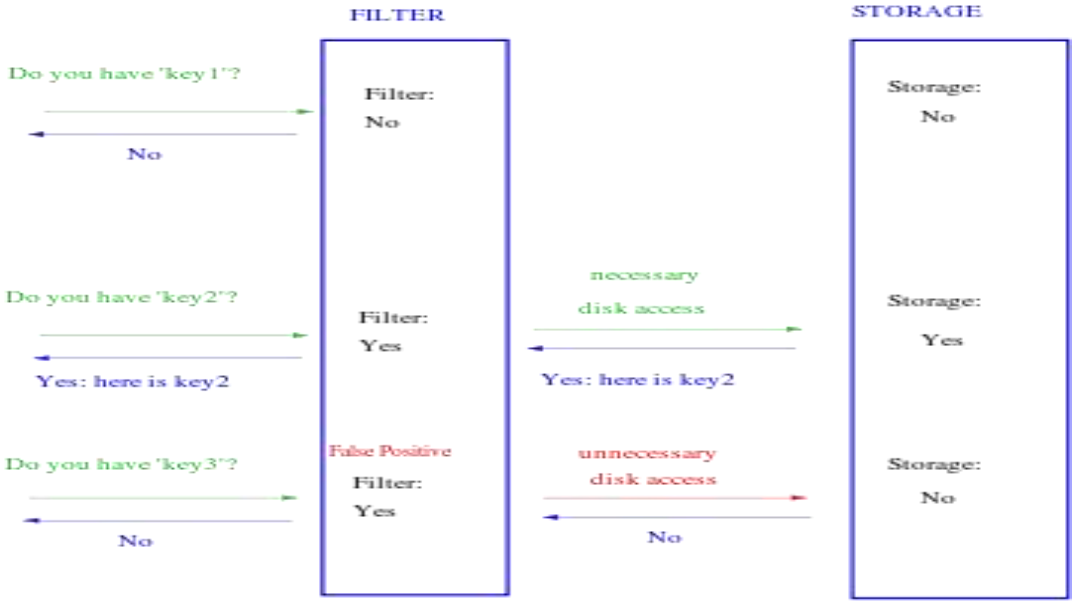
significantly reduce the communication and storage overheads in cloud storage services,

And has prospective applications in our big data-driven society. illegal parties. A variety of data encryption models have been proposed and they are used to encrypt the data before outsourcing to the cloud servers. However, applying these approaches for data encryption usually cause great cost in terms of data utility, which makes traditional data processing methods that are designed for plain text data no longer work well over encrypted data . Secure data deduplication can considerably reduce the communication and storage space overheads in cloud storage services, and has potential applications in our big data-driven society.

Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations; it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge. Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations; it refers to mathematical calculation and algorithmic theme that remodel plaintext into cipher text that may be a non-readable type to unauthorized parties. a spread of knowledge secret writing models have been planned and that they square measure accustomed write in code the data before outsourcing to the cloud servers. However, applying these approaches for encryption typically cause tremendous price in terms of knowledge utility, that makes traditional processing strategies that square measure designed for plaintext knowledge now not work spill encrypted knowledge.

BLOOM FILTER

A Bloom filter may be a space-efficient probabilistic organization, formed by Burton Howard Bloom in 1970, that's wont to check whether or not a component may be a member of a group. False positive matches are attainable, however false negatives aren't – in different words, a question returns either "possibly in set" or "definitely not in set". parts is extra to the set, however not removed (though this could be self-addressed with a "counting" filter); the a lot of parts that are extra to the set, the larger the chance of false positives.



A **Bloom filter** is a space-efficient probabilistic data structure, conceived by Burton Howard Bloom in 1970, that is used to experiment whether an element is a part of a set. False positive matches are possible, but false negatives are not – in new words, a query returns either "possibly in set" or "definitely not in set". Elements can be additional to the set, but not removed (though this can be addressed with a "counting" filter); the more elements that are added to the set, the larger the probability of false positives.

Existing System:

In existing system, Their scheme uses public key encryption to allow keywords to be searchable without revealing data content but investigated the problem for searching over encrypted audit logs. A lot of the early works focused on only keyword searches.

Only just, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords. Other interesting problems, such as the grade of search results and searching with keywords that might contain errors termed unclear keyword search have also been considered.

Existing System Disadvantages:

1. Single keyword search is not smart enough to support advanced queries.
2. Boolean search is unrealistic since it causes high communication cost
3. Need for more search time and difficult to find out.

OBJECTIVE

1. Big data encryption against privacy break.
2. Improve the capability of defending the privacy break.
3. Improve scalability and the time efficiency of query processing

PROPOSED SYSTEM

We gift a phrase search theme that achieves a way quicker reaction time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the collection. We additionally describe modifications to the theme to lower storage value at a tiny low value in latent period and to defend against cloud suppliers with applied mathematics data on keep information. We start by presenting the communication framework and numerous backgrounds as well as connected works. Though phrase searches are a unit processed severallyexploitation our technique, they're generally a specialized perform durin g a keyword search theme, wherever the first perform is to produce conjunctive keyword searches. We describe both basic conjunctive algorithm and phrase search algorithm. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management.

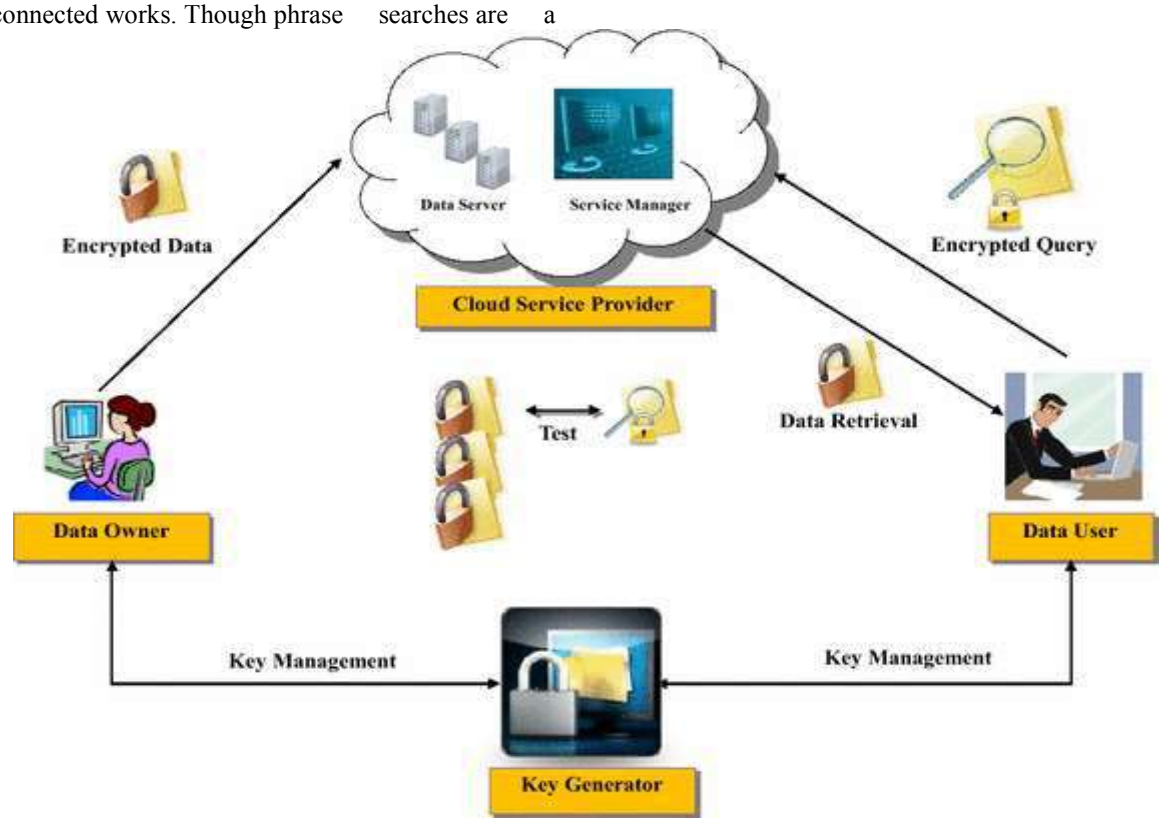
Trapdoor: In Cloud Computing we use this method to encrypt our data with various random possibilities to make our data more secure on cloud. And the similar procedure is follow to decrypt our information in order to reuse our file or any data.

OBJECTIVE

- To reduced the search time
- To enable the multi keyword search over cloud data
- Easy to retrieve the file

SCOPE

- It is scalable, where documents can easily be removed and added to the corpus.
- We also describe modifications to the theme to lower storage cost at a tiny cost in answer time and to defend not in favor of cloud providers with statistical knowledge on stored data.



CONCLUSION

We best owed a phrase search theme supported Bloom alter that's considerably quicker than existing approaches, requiring solely one spherical of communication and Bloom filter verifications. The answer addresses the high process price noted in by reformulating phrase search as n-gram verification instead of a location search or sequent chain verification. Not like our schemes take into account solely the existence of a phrase, omitting any info of its location. Not like our schemes don't need sequent verification, is parallelizable and incorporates a sensible storage demand. In this project, at the time of file uploading on cloud we check file deduplication. We store only unique files on cloud. Using MD5 Algorithm We check file deduplication. File deduplication checking is used for cloud storage management. Our approach is additionally initial the primary to effectively permit phrase search to run severally while not first performing arts a conjunctive keyword search to spot candidate documents. The technique of constructing a Bloom filter index introduced in section allows quick verification of Bloom filters within the same manner as classification. In this procedure are very helpful, so easy to securely store. So it's difficult to hack the file. Therefore, Bloom filter index search technique is used to retrieve the documents extraordinarily very fast.

REFERENCES

1. *D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in In proceedings of Eurocrypt,2004, pp. 506–522.*
2. *B. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, "Building an encrypted and searchable audit log," in Network and Distributed System Security Symposium, 2004.*
3. *M. Ding, F. Gao, Z. Jin, and H. Zhang, "An efficient public key encryption with conjunctive keyword search scheme based on pairings," in IEEE International Conference onNetwork Infrastructure and Digital Content, 2012, pp. 526–530.*
4. *F. Kerschbaum, "Secure conjunctive keyword searches for unstructured text," in International Conference on Network and System Security, 2011, pp. 285–289.*
5. *[R. Carmela, J. Garay, S. Kamara, and R. Ostrovsky, —Searchable symmetric encryption: Improved definitions and efficient constructions,‖in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.*
6. *D. X. Song, D. Wagner, and A. Perrig, —Practical techniques for searches on encrypted data,‖ in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.*
7. *E.-J. Goh et al., —Secure indexes.‖ IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003.*
8. *N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, —Privacy-preserving multi-keyword ranked search over encrypted cloud data,‖ IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.*
9. *Z. Xia, X. Wang, X. Sun, and Q. Wang, —A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data,‖ IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340–352, 2016*