PRIVACY KEYWORD-BASED SEARCH MECHANISMFORDATASTOREDINCLOUD

[1]Bandaru Chandana, [2]Sk Saida

Department of CSE

Dr.Samuel George Institute of Engineering and Technology,Markapur,India

Abstract

Now-a-days, most of the users store their personal and professional data on the cloud. Due to the massive increase inthe storage and computing requirements of users, every time, data is getting transferred to the remote server in larger chunks,without an analysis whether the server on which the data is outsourced, is a trusted server. But the fact is, after outsourcing,users are with a great securityrisk factor that tends to lose the local possession of their large size of data. To maintain theprivacyof personaldocuments stored in cloud environment, itshould getencrypted beforeoutsourcing to thecloud.Afterplacing the data on the cloud, retrieving the same is also a quiet tedious job.In order to retrieve the data, several approaches areavailable in which keyword enabled search of the encrypted data is one of the outstanding techniques. The majority of theseapproaches are limited to handle a single keyword search with its own limitation. To enhance searching in terms of efficiencyand fastness, a multi-key word search technique can be adopted to retrieve a corresponding document from cloud. This paperproposes a survey on a secure search scheme supporting single-keyword or multi-keyword ranked search over encrypted clouddata.

Keywords—Singlekeyword,Multi-keyword search,Rankedsearch,Encryptedclouddata,Security

1.INTRODUCTION

CloudComputingisanincreasingmaturemodelofenterpriseITinfrastructurethatprovidesondemandhighquality applications and services from a shared pool of configuration computing resources. The cloud customers, individuals or enterprises,can outsource their local complex data system into the cloud to avoid the costs of building and maintaining a private storageinfrastructure. The company or organization's private and sensitive information likepersonal files, company records, emails, etcwhich is to be shared among the selected company employees is stored and centralized into cloud server but mostly with an insecurefeeling that anyone may hack these data that may be very risky for that company. Also the data owners and cloud server may not tobeinthesametrusteddomainwhoputtheoutsourcedunencrypteddata,ifany,atrisk;thecloudservermayleakdatainformationtounauthorizedentitiesorevenbehacked.

Cloud enables large group of remote servers to be in a network so as to allow the centralized data repository, and access tothe computer services or resources whenever required. Many users are motivated to outsource their confidential data on to the cloud.As the documents get transferred to the cloud, users don't have physical possession of that data. So as to make sure that the data atcloud side is safer, it has to adapt to the privacy preserving storage, as the cloud server is not a trusted server. To protect dataconfidentiality and unauthorized access to the cloud data, owners are motivated to encrypt their data before it is outsourced to cloud.Toovercomethis issue,thedatastoredincloudstoragedatabaseneeds tobeencryptedbeforesendingtocloudservers forstorage.

Due to the significance of effective storage and computational models available in the cloud, number of cloud service usersis increasing day by day.As a result of this, bulks of information are being getting added into the cloud servers. Insuchenvironment, searching and retrieving the required file is like finding needle onto the sand beach. Sometimes, due toencryptedforms of data on cloud may also force us to have to search

through it, in encrypted form only. This makes the retrieving problem achallenging task. Retrieving files without relevance makes wastage of computational cost and unreliable way to access files. Thus,the use of efficient search technique gains a very high importance nowadays. As user stores the encrypted data at cloud side,traditional searching will not be effective as well. To meet the effective searching on the encrypted cloud data, multi-keyword querycanbeformedsoastogetthetoprelevantdataofuserinterest.

SearchingTechniques:

Therearevarious searchingtechniquesavailable,andtomentionafewareasfollows:

- **Searchable Encryption:** It allows users to securely search complete encrypted data through keywords. This methodsupport only Boolean search, without capturing any relevant data. This approach suffers from two main drawbacks whendirectly applied in the context of Cloud Computing. First one, users who do not necessarily have pre-knowledge of theencrypted cloud data, have to post process every file got, in order, to find ones most matching their interest; anotherdrawback, regularly getting all files containing the queried keyword further incurs unnecessary network traffic, whenretrievemorethanonefiles.

- **Single Keyword Searchable Encryption:** A single keyword searchable encryption schemes usually builds an encryptedsearchable index such that, it's content is hidden to the server, unless it is given appropriate trapdoors generated via secretkey(s). Early work solves secure ranked keyword search which utilizes keyword frequency to rank results instead ofreturning undifferentiated results. However, it only supports single keyword search. Where anyone with public key canwrite to the data stored on server, but only authorized users with private key can search. Traditional single keywordsearchable encryption schemes are usually built in a way by creating an encrypted searchable index. Such indexes contentwill be hidden to the server. The information will be revealed only when the server gives the correct trapdoors that aregenerated via a secret key(s). The main drawback of single keyword-based search is that it is not comfortable enough toexpress complexinformationneeds.

- **Ranked Keyword Search:** Ranked search greatly enhances system usability by returning the matching files in a rankedorderregardingto certainrelevance criteria(eg.keywordfrequency)thus, making one stepclosertowardpracticaldeployment of privacy-preserving data hosting services in the context of cloud computing. To the best of knowledge itgives a legal status for the first time the problem of effective ranked keyword search over encrypted cloud data. Rankedkeyword search strongly provides system usability by returning the matching files in ranked order concerning to certainrelevancecriteria,thusmovingclosetowardsthepracticalactionofprivacypreservingdatapresentings ervicesincloudcomputing. To achieve design goals investigate the statistical measure approach from Information retrieval (IR) and textremoval to insert relevance score of each file during the establishment of searchable index before outsourcing the encryptedfile collection. An IR system allocates a relevance score to each and every document and ranks those documents by thisscore. Relevance score is used to build a secure searchable index to properly protect the sensitive information.Thistechnique enables data users to find the most related information rapidly, rather than burdensome sorting through everymatch in the content collection. Ranked search can also elegantly eliminate unnecessary network traffic by sending backonly the most relevant data. For privacy protection, such ranking function, however, should not leak any keyword relevantinformation. Another one, to improve search result accuracy as well as enhance user searching experience, it is alsoessentialforsuchrankingsystemtosupportmultiplekeywords search.

- **Multi-keyword Searchable encryption:** Over the years, various searchable encryption approaches have been developed toprovide the ability for selectively retrieving the encrypted documents through a keyword search. Typically, these systemsbuild a secure index structure and outsource it along with the encrypted documents to the remote server. Authorized userssubmit their requests as secret trapdoors that are integrated properly with the stored indexing information. The server usesthe received trapdoor to search over the stored index, and retrieves the matching encrypted documents. However, theprevious searchable encryptionschemes are impractical for real world cloud computing scenarios because these systemsaredesignedtohandleeitherasinglekeywordsearchoraBooleansearch.

- **Fuzzy Keyword Searchable Encryption:**Fuzzy keyword search greatly enhances systemusability by returning thematching files when users searching inputs exactly match the predefined keywords or the closest possible matching filesbasedonkeyword similarity semantics,whenexactmatchfails.More specifically,ituseseditdistance to quantifykeywords similarity and develop a novel technique that is a wildcard-based technique, for the construction of fuzzykeyword sets. This technique eliminates the need for enumerating all the fuzzy keywords and the resulted size of the fuzzykeywordsetsissignificantlyreduced.

- **Plaintext Fuzzy Keyword Search:** The importance of fuzzy search has received attention in the context of plaintextsearching in information retrieval community. The problem is addressed in the traditional information access paradigm byallowing user to search without using try-and-see approach for finding relevant information based on approximate stringmatching. At the first glance, it seems possible for one to directly apply these string matching algorithms to the context ofsearchable encryption by computing the trapdoors on a character base within an alphabet. This trivial construction suffersfromthedictionaryandstatisticsattacksandfailstoachievethesearchprivacy.

- **Boolean Keyword Search:** Boolean systems allowed customers to specify their information need using a combination ofBoolean operators AND, OR and NOT. Boolean systems have several disadvantages, for example there are no any featuresof document ranking, and it is very difficult for a customer to make a good search request. Thus, the drawback of existingsystemspecifiestheimportantneedfornewtechniquesthatsupportsearchingflexibility.

2. LITERATURESURVEY

Even though there are various systems existing, this literature survey mainly concentrates on the single keywordbased encryption and multi-keyword based encryptionand also included other searching techniques due to it knownadvantages.

1) **Singlekeywordsearch:**

➢ DeepaliD.Rane et.al,[1]proposedimplementationof the encryptionanddecryption, Secure indexconstructionissuccessfullycompletedwithdesirableperformance.Afterindexconstructionitwillgetcompressedandwillbestoredin
.cfs file format. After firing single-keyword query, user will get all documents that contain the specified keyword. Theadvantages are protects data privacy by encrypting documents before outsourcing, rank based retrievalof the documents,To easily access the encrypted data by multi keyword rank search using keyword index. The Disadvantages of the proposedsystem are single-

keyword search without ranking, Boolean keyword searching without ranking, single-keyword searchwithranking,Rarelysortingoftheresultsi.e.noindexcreationandranking,SingleUsersearch.

➢ C. Wang et al, [2] proposed a secure ranked keyword search technique that utilizes the keyword frequency to rank theresults.The major drawbacks of single keyword search systems with or without ranking most probably won't retrieve therelevantdataanditmayalsocompromisetheprivacy.

➢ Y.-C. Chang et al, [3] proposed similar "index" approaches, in which a single encrypted hash table index was built for theentirefilecollection.

➢ D. Song, D. Wagner et al, [4] proposed a searchable encryption, in which each and every word in the considered documentwas encryptedautonomouslyundertwo-layeredencryptionconstruction.

3.Multi-keywordsearch:

➢ Zhihua Xia et.al,[5] proposed a secure, efficient and dynamic search scheme, which supports not only the accurate multi-keyword ranked search but also the dynamic deletion and insertion of documents. They construct a special keywordbalanced binary tree as the index, and proposed a "Greedy Depth-first Search" algorithm to obtain better efficiency thanlinear search. In addition, the parallel search process can be carried out to further reduce the time cost. The security of thescheme is protected against two threat models by using the secure KNN algorithm. Experimental results demonstrate theefficiency of proposed scheme. The Advantages of the proposed system are searchable encryptionschemes enable theclient to store the encrypted data to the cloud and execute keyword search over cipher text domain and a secure tree-basedsearch scheme over the encrypted cloud data, which supports multi-keyword ranked search and dynamic operation on thedocument collection. The disadvantages are the cloud service providers (CSPs) that keep the data forusers may accessusers sensitive information without authorization. A general approach to protect the data confidentiality is to encrypt thedatabeforeoutsourcing.However, thiswillcauseahugecostinterms ofdatausability.

➢ Bing Wang et.al, [6] proposed a novel construction of a public key searchable encryption scheme based on inverted index.This scheme overcomes the one-time-only search limitation in the previous schemes. The disadvantages of the proposedsystem are first of all, the keyword privacy is compromised once a keyword is searched. As a result, the index must berebuilt for the keyword once it has been searched. Such solution is counterproductive due to the high overhead suffered.Secondly, the existing inverted index based searchable schemes do not support conjunctive multi-keyword search, which isthe most common form of queries now a days. The advantages are explore the problem of building a searchable encryptionscheme based on the inverted index, Achieve secure and private matching between the query trapdoor and the secure index,Design a novel trapdoor generation algorithm so that the query related inverted lists are combined together secretly withoutlettingthecloudserverknowwhichinvertedlistsareretrieved.

➢ YanzhiRen et.al, [7] proposed a light-weight search approach that supports efficient multi-keyword ranked search in cloudcomputing system. The basic scheme employs the polynomial function to hide the encrypted keyword and search patternsfor efficient multi-keyword ranked search. Then improve the basic scheme and propose a privacy-preserving scheme whichutilizes the secure inner product method for protecting the privacy of the searched multi-keywords. The advantage of theproposed

system is it analyzes the privacy guarantee of the proposed scheme and conduct extensive experiments based onthereal-worlddataset. Thedisadvantageisthereisachanceofleakageofdataincloud.

➢ Hongwei Li et.al, [8] proposed a multi-keyword ranked search scheme to enable accurate, efficient and secure search overencryptedmobileclouddata.Securityanalysishavedemonstratedthatproposedschemecaneffectivelyac hieveconfidentiality of documents and index, trapdoorprivacy,trapdoorunlinkability,and concealing access patternof thesearch user. The advantages Constructs an efficient index to improve the search efficiency. And it solves the trapdoorunlinkability problem. It also achieve enhanced efficiency in terms of functionality and search efficiency compared withexistingproposals.

➢ Mikhail Strizhov et.al, [9] proposedasearchable encryption technique that enables secure searches over encrypted datastored on remote servers. They define and solve the problem of multi-keyword ranked search over encrypted cloud data. Inparticular,theypresentanefficientsimilaritysearchableencryptionschemethatsupportsmulti-keywordsemantics.Thesolution is based on two building blocks: Term Frequency Inverse Document Frequency (TF-IDF) measurement and ring-LWE-basedvariantofhomomorphiccryptosystem.TheAdvantagesofthissystemisitreturns thematchingdataitems inarankedorderedmanner.TheDisadvantageisintraditionalsystemitsupportsonlysinglekeywordsearch.

4. Othersearchingtechniques:

➢ E.-J. Goh et al, [10] proposed a technique that uses Bloom filters in order to construct the indexes for the data files. Bloomfilter containing trapdoors (for each file) of all distinct words is built up and stored on the server. For searching a particularword, the user must generate the search request by computing the trapdoorof the word and sends it to the server. Theserver upon receiving the request performs tests to check if any Bloom filter holds the trapdoor of the query word and if so,itreturnsthecorrespondingfileidentifiers.

➢ Jun Zhou et.al, [11] proposed a more efficient verifiable outsourced computation of encrypted data EVOC from any one-way trapdoor function is proposed by combining a newly devised privacy-preserving data aggregation supporting bothaddition and multiplication operations with Yao's Garbled Circuit. The advantage is it proves the security of the proposedefficientprivacy-preservingdataaggregationscheme.

➢ Fanyu Bu et.al, [12] proposed a privacy preserving backpropagation algorithm based on the BGV encryption scheme oncloud. One property of the proposed algorithm is to apply the BGV encryption scheme to the back-propagation algorithmfor preventing the disclose of private data with cloud computing. The advantages: The proposed algorithm improves theefficiency of back-propagation learning by offloading the expensive operations on the cloud. It also preventsthe disclosureofprivate data, using full homomorphic encryption scheme to encrypt the source data. The disadvantage is sensitive data iseasilydisclosedduringtheprocessofthecomputationonthecloud.

➢ Joseph K et.al, [13] proposed an infrastructure for secure sharing and searching for real-time video data. It is particularlysuitable for mobile users by deploying 5G technology and a cloud computing platform. The security is guaranteed even ifthe cloudserverishackedsince dataconfidentiality isnow

protectedby cryptographic encryptionalgorithms. Theadvantage of the proposed system is the infrastructure security is guaranteed even if the cloud server is hacked. Thedisadvantage is There are some existing platforms for sharing real time video, they may not be able to achieve secure fine-grainedsharingandsecuresearchingsimultaneously.

- Zhangjie Fu et.al, [14] proposed an efficient verifiable keyword-based semantic search scheme. Comparing to most of theexisting searchable encryption schemes, the proposed scheme is more practical and flexible, better suiting user's differentsearch intentions. Moreover, the proposed scheme protects data privacy and supports verifiable search ability, in thepresence of the semihonest server in the cloud computing environment. The Advantages: Improves the flexibility andsupportverificationofsearchresults.Alsoitprovidesverifiablesearchabilitywithdataprivacypreserving. Thedisadvantage is The trivialsolutionof downloading the whole encrypted datafirst and thendecrypting it locally isobviouslyimpractical, duetothehugebandwidthandcomputationburden.

- Jin Li et al, [15] proposed a new searching technique that is fuzzy keyword search. They focused on enabling effective yetprivacy preserving fuzzy keyword search in Cloud Computing. To the best of knowledge, they formalize for the first timetheproblemofeffectivefuzzykeywordsearchoverencryptedcloud datawhilemaintainingkeywordprivacy.

- D. Boneh, G. D. Crescenzo et.al, [16] proposed a searchable encryption, in which anyone who has the public key can writeto the data stored in server but with a restriction that only authorized users with private key can search.The major demeritsof using public key are that it's very expensive in terms of computation.Additionally, in the public key setting, the privacyof keyword may possibly not be protected as the serverpossesses the ability to encrypt any keywordwith the public key.As a result of this it can be used to receive the trapdoor in order to evaluate the cipher text. For achieving more efficientsearch,Curtmolaetal. singlekeywordsearch.

5. CONCLUSION

This paper proposes brief literature survey on single keyword based searching, multi-keyword based searching and many othersearching techniques in cloud based environment. From the survey, the multi-keyword search technique sounds to be more efficientthan other available searching technique. Many search schemes over encrypted data, supports multi-keyword query and similarityrankingsimultaneouslyfordataretrievalincloudcomputing.

6. REFERENCES

[1] Deepali D. Rane andDr.V.R.Ghorpade" Multi-User Multi-Keyword Privacy Preserving Ranked Based Search Over EncryptedCloudData"InternationalConferenceonPervasiveComputing(ICPC), 2015.

[2] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in Proc. of ICDCS'10,2010.

[3] Y.-C.ChangandM.Mitzenmacher,"Privacypreservingkeyword searchesonremoteencrypteddata,"inProc.ofACNS,2005.

[4] D.Song,D.Wagner,and A.Perrig,"Practicaltechniquesforsearchesonencrypted data,"inProc.ofS&P,2000.

[5] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang,

Member, IEEE "A Secureand Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data" IEEE TRANSACTIONS ON PARALLEL ANDDISTRIBUTED SYSTEMS,VOL.,NO.1,2015.
[6[BingWang,WeiSong,WenjingLou,andY.Thomas Hou"InvertedIndexBasedMulti-KeywordPublic-keySearchableEncryption with Strong Privacy Guarantee" IEEE Conference on Computer Communications (INFOCOM), 2015.[7]YanzhiRen,YingyingChen,JieYang,BinXie"Privacy-preservingRankedMulti-KeywordSearchLeveragingPolynomialFunctioninCloudComputing"GlobecomCommunicationandInformation SystemSecuritySymposium2014.

[8] Hongwei Li, DongxiaoLiu,Yuanshun Dai, Tom H. Luan, And Xuemin (Sherman) Shen "Enabling Efficient   Multi-KeywordRankedSearchOverEncryptedMobileCloudDataThroughBlindStorage",December2014.

[9] Mikhail Strizhov and Indrajit Ray "Multi-keyword Similarity Search Over Encrypted Cloud Data" International Conference onPervasiveComputing(ICPC), 2012.

[10] E.-J. Goh, " Bloom filters in order to construct the indexes for the data files" IEEE Conference on Computer Communications2016.

[11] Jun Zhou, Zhenfu Cao, Xiaolei Dong and Xiaodong Lin "More Efficient Verifiable Outsourced Computation from Any One-wayTrapdoorFunction"IEEEICC-CommunicationandInformationSystems SecuritySymposium,2015.

[12] Fanyu Bu, Yu Ma, Zhikui Chen and Han Xu "Privacy Preserving Back-Propagation Based on BGV on Cloud" 2015 IEEE 17thInternational Conference on High Performance Computing and Communications (HPCC), 2015 IEEE 7th International SymposiumonCyberspaceSafetyandSecurity(CSS),and 2015IEEE12thInternationalConfonEmbedded Softwareand Systems(ICESS).

[13] JosephK,"SecureSharingandSearchingforReal-TimeVideoDatainMobileCloud"2015.

[14] Zhangjie Fu, *Member,* IEEE, JiangangShu, Xingming Sun, and Nigel Linge "Verifiable Keyword-based Semantic Search overEncryptedCloudData" IEEETransactionsonConsumerElectronics,Vol. 60,No.4,November2014.

[15] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "Fuzzy keyword search over encrypted data in cloud computing," inProc.ofIEEEINFOCOM'10Mini-Conference,SanDiego,CA,USA,March2010.

[16] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryptionwith keyword search," in Proc. ofEUROCRYPT,IEEEConferenceonComputerCommunications2004.