TKSE: Trustworthy keyword search over encrypted data with two side verifiability via block chain

Rohini M, EPCET, Bangalore, India rohiniamlu998@gmail.com
Dr. Nanda ashwin, Asst Professor, EPCET, Bangalore, India, nandaashwin7@gmail.com
Sneha C, EPCET, Bangalore, India, snehachandra9845@gmail.com
Sushmitha N, EPCET, Bangalore, India, nsushmitha34@gmail.com
Yashaswini M N, EPCET, Bangalore, India, yashaswinimnn@gmail.com@gmail.com

Abstract: The cloud computing is very attractive model, cloud computing makes it possible for resource constrained users to enjoy cost-effective and flexible resources of various things. The untrustworthiness of cloud servers and the data privacy of users are necessary to encrypt the data before uploading it to the cloud. Encryption may cause a serious of problems , like as: how can the user searches the uploaded data? How users can secure there search results to avoid malicious cloud servers ? How to enable server-side verifiability of uploading data to check malicious data owner? in this paper, we introduce TKSE, a Trustworthy Keyword Search scheme over Encrypted data without any third party, We use a digital signature for encryption is used to search the data that has been uploaded to the cloud and it again checks the searched data which is retrieved by the cloud. TKSE assumes server-side verifiability which can saves honest cloud servers from being framed by intending to cause data owners in the data storage phase. The blockchain technology and hash functions are used here to avoid the thirdparties even if the user or cloud isdangerous.

Keywords: — ***Blockchain, Cloud computing, Searchable encryption, Server side verifiability, User side Verifiability***

I. Introduction

CLOUD computing has been considered as a newmmodel of enterprise IT infrastructure, which can organize huge resource of computing, storage and applications, and enable users to enjoy ubiquitous, convenient and needed network access to a shared pool of configurable computing resources with great efficiency and minimal economic overhead. Attracted by these appealing features, both individuals and enterprises are motivated to outsource their data to the cloud, instead of purchasing software and hardware to manage the data themselves. Despite of the various advantages of cloud services, outsourcing sensitive information (such as e-mails, personal health records, company finance data, government documents, etc.) to remote servers brings privacy con-cerns. The cloud service providers (CSPs) that store the data for users can access user's sensitive information without authorization. A near approach to preserve the data confidentiality is to encrypt the data before outsourcing. However, this will cause a huge cost in terms of data usability. For example, the existing techniques based on keyword information retrieval that are widely used on the plaintext data cannot be directly applied on the encrypted data. Downloading all the data through the cloud and decoded data locally that are evidently impractical. In order to address the above problem, researchers have planned some general purpose solutions with fully similar encryption or oblivious RAMs. However, these methods are not practical due to their high computational overhead for both the cloud sever and user.

On the contradictory, more practical special-purpose solutions, such as searchable encryption (SE) schemes have made specific contributions in terms of efficiency, functionality and security. Searchable encryption schemes authorize the client to store the encrypted data to the cloud and execute keyword search over cipher text domain. Recently, abundant works has been proposed under different threat models to achieve various search functionality, such as single keyword search, similarity search, multi-keyword Boolean search, ranked search, multi-keyword ranked search, etc. Among them, multi-keyword, ranked search to reach more and more notice for its practical applicability. Recently, some dynamic schemes are projected to support inserting and deleting operations on document assortment. These are significant works because it is highly possible that the data owners need to update their data on the cloud server. But few of the dynamic schemes support economical multi-keyword hierarchy search.

DOI : 10.18231/2454-9150.2019.0575
© 2019, IJREAM All Rights Reserved.

II. LITERATURE SURVEY

[1] Security in cloud computing: Opportunities and challenges

The cloud computing exhibits, exceptional potential to produce price effective, simple to man-age, elastic, and powerful resources on the fly, over the net. The cloud computing, upsurges the capabilities of the hardware resources by optimum and shared utilization. The higher than mentioned options encourage the organizations and individual users to shift their applications and services to the cloud. Even the vital infrastructure, for example, power generation and distribution plants are being migrated to the cloud computing paradigm. However, the services provided by third-party cloud service suppliers entail extra security threats. The migration of user's assets (data, applications, etc.) outside the executive management in an exceedingly shared surroundings where varied users are collocated escalates the safety considerations. This survey details the safety problems that arise because of the terribly nature of cloud computing. Moreover, the survey presents the recent solutions given within the literature to counter the safety problems. Furthermore, a short read of security vulnerabilities within the mobile cloud computing also are highlighted. In the end, the discussion on the open problems and future analysis directions is additionally given.

[2] A Trusted Solution for Secure Outsourced Data Using Modified Key Policy Attribute Based Encryption

Cloud computing is a potential and emerging technology for next generation of computer applications. The secured data sharing methods provides security between data owner and user with the user's attribution. The obstacle and hurdles toward the rapid growth of cloud system affected by some major issues such as data security, privacy, and data sharing. There are so many researchers have introduced different techniques for data defense also to attain highest level of data security in the cloud computing system. However, those have some issues, thus we required to solve in more effective way. Furthermore, we also need to guarantee privacy for the outsourced data, and bring no additional local search burden to user. This paper proposed an efficient file modified key policy attribute-based encryption scheme. The layered access structure has been integrated into a single access structure, and the stored files encrypted by integrated access structure. Then the secret data has encrypted by data owner using modified KP-ABE Encryption scheme. The cipher text elements associated with attributes may be shared by the files. Therefore, both cipher text storage time and cost of encryption stored. Thus the proposed system gives the greater performance and security to the distributed data sharing system.

[3] Practical Techniques for Searches on Encrypted Data

It is fascinating to store data on data storage servers like mail servers and file servers in encrypted type to scale back security and privacy risks. But this typically implies that one needs to sacrifice practicality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the question while not loss of information confidentiality . In this paper, we describe our cryptographic schemes for the problem of searching on encrypted data and provide proofs of security for the resulting crypto systems. Our techniques have a number of crucial advantages. They are demonstrably secure: they supply demonstrable secrecy for encoding, in the sense that the untrusted server cannot learn anything about the plaintext when only given the ciphertext; they provide question isolation for searches, meaning that the untrusted server cannot learn anything more about the plaintext than the search result; they provide controlled searching, so that the untrusted server cannot explore for associate capricious word while not the user's authorization; they also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server .

[4] Public Key Encryption with keyword Search

We study the matter of looking out on data that's encrypted using a user public key system. Consider user Bob who sends email to user Alice encrypted under Alice's public key . An email gateway wants to test whether the email contains the keyword \urgent" so that it could route the email accordingly. Alice, on the opposite hand doesn't want to allow the entrance the flexibility to rewrite all her messages. We dene and construct a mechanism that enables Alice to provide a key to the gateway that enables the gateway to test whether the word \urgent" is a keyword in the e-mail while not learning anything the rest concerning the email. We discuss with this mechanism as Public Key Encoding with keyword Search. As another example, consider a mail server that stores various messages publicly encrypted for Alice by others. Using our mechanism Alice will send the mail server a key which will alter the server to spot all messages containing some specic keyword, but learn nothing else. We dene the concept of public key encryption with keyword search and give several constructions.

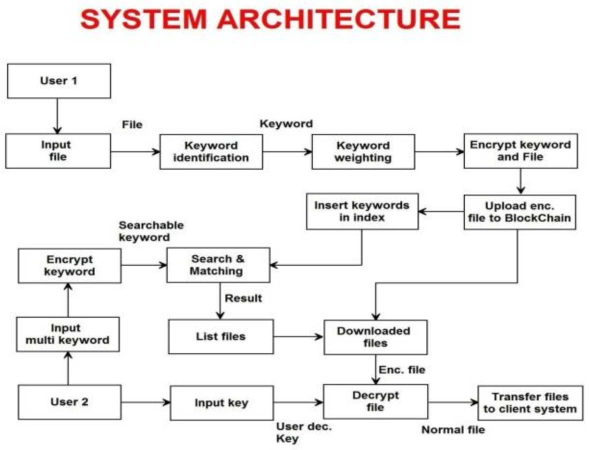
[5] UC-Secure Searchable Symmetric Encryption

For searchable radial encoding schemes (or symmetric-key encoding with keyword search), the safety against passive adversaries (i.e. privacy) has been mainly considered so far. In this paper, werst dene its security against active adversaries (i.e. reliability as well as privacy). We next formulate its UC-security. We then prove that the UC-security against non-adaptive adversaries is equivalent to our dentition of privacy and reliability. We further present an ecient construction which satises our security dentition (hence UC-security).

DOI : 10.18231/2454-9150.2019.0575 © 2019, IJREAM All Rights Reserved.

III. SYSTEM ARCHITECTURE

System Architecture design-identifies the overall hypermedia structure for the WebApp. Architecture style is tied to the goals establish for a WebApp, the content to be conferred, the users who will visit, and the navigation philosophy that has been established. Content design, focuses on the style within which content objects and structured for presentation and navigation. WebApp architecture, addresses the manner in which the application is structure to manage user interaction, handle internal processing tasks, effect navigation, and present content. WebApp architecture is defined within the context of the development environment in which the application is to be implemented.



A. Existing system or Methodology

- Case 1- Malicious CSP may try to learn information on secret keys and original documents of User from trapdoor and cipher text data.
- Case2-MaliciousCSPmayforgesearchresultsorreturn partial results.
- Case3-Malicious User may defraud CSP for compensation by outsourcing invalid cipher text data and index to CSP in the data storage phase.
- Case4-Malicious CSP aims to obtain search fees from User without providing search results satisfying the requirements of User.
- Case5-Malicious User wants to get valid search results from CSP without paying the search fee. Considering the above attacks, the design goals of TKSE are as follows.
- Privacy-In TKSE, it is impossible for CSP to learn any information on secret keys and original documents. This goal is against the attack in Case 1.
- User-side Verifiability- It is the same as the property reliability.This goal is against the attack in Case2. That is, CSP cannot forge valid search results or return partial results.
- Server-side Verifiability. This goal is against the attack in Case 3 and it can protect CSP from being framed by malicious User.
- Case 5. Specifically, fairness ensures that User can get valid search results only if corresponding each fee are paid to CSP. On the other hand, CSP obtains search fees from User only if satisfactory search results are returned. In addition, considering that blockchain technologies are used in TKSE, the following attractive properties should be achieved.

Compatibility- TKSE should be compatible with mainstream blockchains such as Bitcoin and Ethereum. Note that if TKSE is compatible with the Bitcoin blockchain, then it is compatible with Ethereum blockchain.

B. Proposed System

TKSE is composed of four phases including the initialization phase, the data storage phase, the data search phase and the user claim phase. The first three phases are compulsory and the user claim phase is performed by User when CSP is malicious. The definition of TKSE is as follows:

1. INITIALIZATION User and CSP initialize parameters to be used in the subsequent phases, such as their own secret keys and unredeemed transactions on the block chain.

2. DATA STORAGE PHASE The outsourcing of data storage is implemented in this phase. Four procedures, data encryption, index generation, storage enforcement and storage confirmation, are sequentially performed as follows:

- Data Encryption: For confidentiality, User encryptsD to getC which is outsourced to CSP.
- Index Generation: User generates I and sends it together with C to CSP.
- Storage Enforcement: CSP firstly checks and stores IandC. Then, CSP generates a digital signature according to C and stores the signature on the blockchain. Finally, CSP sends such information to User that helps User to get the signature from the blockchain. • Storage Confirmation: Upon getting the signature from the blockchain, User thinks that C has been stored by CSP.

3. DATA SEARCH PHASE

In this phase, CSP searches his/her storage according to the requirements of User. It can be ensured that CSP earns search fees from User if and only if User obtains valid search results. Four sequential sub-phases are performed, including search request, search commitment, payment commitment, and verification and payment.

4. USER CLAIM PHASE

Only if CSP fails to prove that the search result meets the requirements of User before the given time, TKSE comes to User Claim Phase. It is used by User to claim enough deposits of CSP no matter how CSP behaves.

C. ADVANTAGE OF PROPOSED SYSTEM:

- Dynamic
- Search Efficiency
- Keyword privacy

D.SYSTEM REQUIREMENT SPECIFICATION

Software requirement Specification is a fundamental document, which forms the foundation of the software development process. It not only lists the requirements of a system but also has a description of its major feature. An SRS is basically an organization's understanding (in writing) of a customer or potential client's system requirements and dependencies at a particular point in time (usually) prior to any actual design or development work. It's a two-way insurance policy that assures that both the client and the organization understand the other's requirements from that perspective at a given point in time. The SRS also functions as a blueprint for completing a project with as little cost growth as possible. The SRS is often referred to as the "parent" document because all subsequent project management documents, such as design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans, are related to it. It is important to note that an SRS contains functional and nonfunctional requirements only; it doesn't offer design suggestions, possible solutions to technology or business issues, or any other information other than what the development team understands the customer's system requirements to be.

E. FUNCTIONAL REQUIREMENT

A Software Requirement Specification (SRS) is basically an organization's understanding of a customer or potential client's system requirements and dependencies at a particular point prior to any actual design or development work. The information gathered during the analysis is translated into a document that defines a sets of requirements. The SRS gives the complete details of the services that the system should provide and also the curb under which the system should operate Generally, the SRS is a document that describes what the proposed software should do without explaining the how the software do it. It's a two-way document detailing the terms and conditions that assures that both the client and the organization understand the other's requirements from that perspective at a given point in time. The SRS also acts as a outline for completing a project with as little cost growth as possible. Because of all the next project management documents, such as design specifications, statements of work, software architecture specifications, testing and validation plans, and documentation plans, are related to it, So SRS is often referred to as the "parent" document. Requirement could be a condition or capability to that system should change .Requirement Management could be a systematic approach towards eliciting, organizing and documenting the requirements of the system clearly along with the applicable attributes. The elusive difficulties of needs don't need to be continuously obvious and might come back from any range of sources.

Functional demand defines a performance of a package and the way the system should behave once given with specific inputs or conditions.These may include calculations, data manipulation and processing and other specific functionality.

In this system following are the functional requirements:-

•Collect the lung disease data set of patients.
•Train the Patient Data.
•Predict the Disease for patient.

F. NON-FUNCTIONAL REQUIREMENT

Non-functional requirements are the requirements which are not directly concerned with the specific function delivered by the system. They specify the standards which will be wont to decide the operation of a system instead of specific behaviors .They may relate to emerging system properties like dependableness, response time and store occupancy Non-functional wants arise through the user desires, because of budget constraints, organizational policies, and the need for interoperability with other software and hardware systems or because of external type of the system which are observed during run time, whereas evolution quality involves testability, maintainability, extensibility or scalability.

IV. RESOURCE REQUIREMENT

DOI : 10.18231/2454-9150.2019.0575 © 2019, IJREAM All Rights Reserved.

A. THE JAVA PROGRAMMING LANGUAGE:

The Java programming language could be a application-oriented language which will be characterized by all of the subsequent buzzwords::

- Simple
- Architecture neutral
- Object oriented
- Portable
- Distributed
- High performance
- Interpreted
- Multithreaded
- Robust
- Dynamic
- Secure

With most programming languages, you either compile or interpret a program so that you can run it on your computer. The Java programming language is uncommon therein a program is each compiled and understood. With the compiler, first you translate a program into an intermediate language called Java byte codes the platform-independent codes interpreted by the interpreter on the Java platform The interpreter parses and runs every Java computer memory unit code instruction on the pc . Compilation happens simply once; interpretation happens anytime the program is dead.

The following figure illustrates how this works.

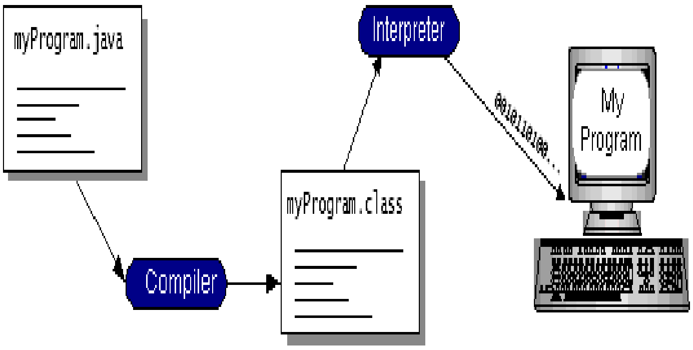


Fig: Java Virtual Machine

You can consider Java computer memory unit codes because the computer code directions for the Java Virtual Machine (Java VM).

Every Java interpreter, whether or not it's a development tool or an internet browser which will run applets, is Associate in Nursing implementation of the Java VM.

Java computer memory unit codes facilitate create "write once, run anywhere" possible.

You can compile your program into computer memory unit codes on any platform that contains a Java compiler.

The computer memory unit codes will then be run on any implementation of the Java VM.

That means that as long as a transferable portable a Java VM, the same program written in the Java programming language can run on Windows 2000, a Solaris workstation, or on an iMac.

computer moveable laptop contains

B. THE JAVA PLATFORM:

A platform is that the hardware or computer code surroundings within which a program runs.We have already mentioned a number of the foremost in style platforms like windows 2000, linux, solaris, and MacOS. Most platforms will be delineated as a mix of the package and hardware.

The Java platform differs from most differ that runs on highs from most different platforms in it's software only platform

The Java platform differs from most different platforms in this it's a software-only platform that runs on high of different hardware-based platform

The Java platform has two components:

- The Java Virtual Machine (Java VM)
- The Java Application Programming Interface (Java API)

The Java platform gives you the following features:

- **The Essentials**: Objects, strings, threads, numbers, input and output, data structures, system properties, date and time, and so on.
- **Applets**: The set of conventions used by applets.
- **Networking**: URLs, TCP (Transmission Control Protocol), UDP (User Data gram Protocol) sockets, and IP (Internet Protocol) addresses.
- **Internationalization**: Help for writing programs that can be localized for users worldwide.Programs will mechanically adapt to specific locales and be displayed within the acceptable language..
- **Security**: Both low level and high level, including electronic signatures, public and private key management, access control, and certificates.
- **Software components**: Known as JavaBeans ^{TM}, can plug into existing component architectures.
- **Object serialization**: Allows lightweight persistence and communication via Remote Method Invocation (RMI).
- **Java Database Connectivity (JDBC)**: Provides uniform access to a wide range of relational databases.

DOI : 10.18231/2454-9150.2019.0575
© 2019, IJREAM All Rights Reserved.

The Java platform in to boot has arthropod genus for second and 3D graphics, accessibility, servers, collaboration, telephony, speech, animation, and more.

C.HTML

HTML, associate initialism of machine-readable text terminology, is the predominant markup language for web pages. It provides a way to explain the structure of text-based information during a document — by denoting sure text as headings, paragraphs, lists, and so on — and to supplement that text with interactive forms, embedded images, and other objects. HTML is written within the sort of labels (known as tags), surrounded by angle brackets. HTML also can describe, to some degree, the appearance and semantics of a document, and can include embedded scripting language code which can affect the behavior of web browsers and other HTML processors.

HTML is additionally usually accustomed ask content of the MIME sort text/html or perhaps more broadly as a generic term for HTML whether or not in its XML-descended kind (such as XHTML 1.0 and later) or its kind descended directly from markup language

Hypertext Markup Language (HTML), the languages of the World Wide Web (WWW), allows users to produces Web pages that include text, graphics and pointer to other Web pages (Hyperlinks).

HTML isn't a artificial language however it's an application of ISO Normal 8879, SGML (Standard Generalized Markup Language), but specialized to hypertext and adapted to the Web. The idea behind machine-readable text is that rather than reading text in rigid linear structure, we can easily jump from one point to another point. We can navigate through the data supported based on our interest and preference. A nomenclature is just a series of parts, each delimited with special characters that define how text or other items enclosed within the elements should be displayed. Hyperlinks square measure underline or stressed works that load to different documents or some parts of an equivalent document.

HTML may be accustoned show any sort of document on the host laptop, which can be geographically at a different location. It is a flexible language and may be used on any platform or desktop.

HTML provides tags (special codes) to create document look enagaging. HTML tags are not case-sensitive. Using graphics, fonts, totally different sizes, color, etc., can enhance the presentation of the document. Anything that's not a tag is an element of the document itself.

V. ALGORITHM

A.FOR BUILDING INDEX

BuildIndex(K,D,C,W)

Input: secret key K, document setD, ciphertext data setC, keyword setW

Output : index I

```
 1  for i ∈ [num] do
 2  I(i) = (dummy,sigU(i k H(dummy)))
 3  end
 4  for ω ∈W do
 5      SupposeDω = {Ds1,Ds2,...,Dsm}.
 6      for j ∈ [m] do
 7          addrω,j = π(0,ω,j)
 8          tagω,j = sigU(addrω,j k H(Csj)),
 9          I(addrω,j) = (sj,tagω,j)
10      end
11  end
12  for Dk ∈Ddo
13      Suppose k has already appeared Nk times inI.
14      if Nk 6= 2`w then
15          for 1 ≤ r ≤ 2`w −Nk do
16              addrr,k = π(1,r,k)
17              tagr,k = sigU(addrr,k k H(Ck))
18              I(addrr,k) = (k,tagr,k)
19          end
20      end
21      Now, the index k appears 2`w times inI.
22  end
23  return I
```

B.SEARCH ALGORITHM

Search(I,C,Tω) → (Cω,tagω)

Input: indexI, ciphertext data setC, trapdoor Tω

Output: ciphertext data setCω and tag set tagω with respect to ω

```
1 LetCω = ∅.
2 Parse Tω = (addrω,1,addrω,2,...,addrω,N)
3 for j ∈ [N] do
4 parseI(addrω,j) = (sj,tagω,j)
5 if sj
6= dummy then 6 setCω = Cω ∪Csj
7 end
8 end
9 Set tagω = (tagω,1,tagω,2,...,tagω,N).
10  eturnCω and tagω
```

C.VERIFY ALGORITHM

Verify (Tω,Cω,tagω) →1 or 0

Input: trapdoor Tω, ciphertext data setCω and tag set tagω related to ω

Output: accept(1) or reject(0)

```
1.Parse Tω = (addrω,1,addrω,2,...,addrω,N)
 2 ParseCω = {Csi}i∈Iω.
2  Parse tagω = (tagω,1,tagω,2,...,tagω,N).
3  for j ∈ [N] do
```

International Journal for Research in Engineering Application & Management IJREAM

DOI : 10.18231/2454-9150.2019.0575

© 2019, IJREAM All Rights Reserved.

```
  if $j \in I\omega$ then
    set $h* j = H(Csj)$
  else
    set $h* j = H(dummy)$
  end
  if$verU(addr\omega,j\ k\ h*\ j,tag\omega,j) = false$then
      return 0
    end
  end
  return 1
```

VI. CONCLUSION

Considering the sensible drawback of privacy knowledge sharing system supported public cloud storage which needs an data owner to distribute an oversized variety of keys to users to enable them to access his/her documents, we for the first time propose the concept of Secure and Dynamic multi keyword ranked search over encrypted data scheme. Both analysis and evaluation results confirm that our work can provide an effective solution to building practical data sharing system based on public cloud storage.

REFERENCES

[1] M. Ali, S. U. Khan, and A. V. Vasilakos, “Security in cloud computing”: Opportunities and challenges," Inf. Sci Jun 2015.

[2] Z. Wan, J. Liu, and R. H. Deng, ``HASBE: A hierarchical attribute-based solution for exible and scalable access control in cloud computing," *IEEE* Trans. Inf. Forensics Security, vol. 7, no. 2, pp. 743_754, Apr. 2012.

[3] D. X. Song, D. Wagner, and A. Perrig, ``Practical techniques for searches on encrypted data," in *Proc. IEEE Symp, Secur. Privacy*, May 2000, pp. 44_55.

[4] D. Boneh,G.Di Crescenzo, R. Ostrovsky, and G. Persiano, ``Public key encryption with keyword search," in *Eurocrypt*, vol. 3027. Berlin, Germany: Springer, 2004, pp. 506_522.

[5] K. Kurosawa and Y. Ohtaki, ``UC-secure searchable symmetric encryption”, in *Financial Cryptography*, vol. 7397. Berlin, Germany: Springer, 2012, pp. 285_298.

DOI : 10.18231/2454-9150.2019.0575
© 2019, IJREAM All Rights Reserved.