VP Search: Achieving Verifiability of search result and privacy protection for outsourced data at the same time.

Ravindra Hyalij [1]
[1]P.G. Student, Department of Computer Engineering, SKN Sinhgad Institute of Technology & Science, Lonavala, Pune, Maharashtra, India

ABSTRACT

Cloud computing offers flexible computation and resources for storage purpose, but here user poses challenges on verifiability of computations and data privacy. Because of data privacy reason some data owner away from cloud facility. Proposed system resolve this problem by providing facility of verifiability for privacy preserving multi-keyword search over outsourced documents. In this system integration of two technique use first is homomorphic MAC technique and second one is privacy-preserving multi-keyword search. The proposed scheme enables the client to verify search results efficiently without storing a local copy of the outsourced data.

Keyword: *Decryption, Encryption, Multi keyword Search technique, Private Cloud, Public Cloud*

1. INTRODUCTION

Cloud computing provide flexible computation to achieve computation in less time and storage space for storing huge amount of data. A client can outsource his data to the cloud server in encrypted form and later access the data with other devices by using single keyword and multi keyword search technique from anywhere. The client can further ask the cloud server to perform some computation over his data on his behalf. For access outsource data client have to create trapdoor. For creating trapdoor client register their identity token on data owner side then owner give him decryption key of data. But, here user faces some challenges like verifiability of computations and data privacy. In this study the focus is on verifiability for privacy-preserving multi-keyword search over outsourced documents.

2. EXISTING SYSTEM

In existing system, verifiable computation schemes provide facility to transfer certain computing task on the cloud with intent to minimize computation effort of the client, but not storage or communication cost. They require the client and the server to interactively authenticate the computation result. Moreover, the client needs to have a copy of the outsourced data, and the data over which computation is verified cannot be changed in the future. Existing system supports homomorphic MAC (Message Authentication Code) scheme in this scheme homomorphic message authenticators allow to validate computation on previously signed data. The holder of a dataset$\{m_1,...,m_l\}$ uses her secret key *sk* to produce corresponding tags $(\sigma_1,...,\sigma_l)$ and stores the authenticated dataset on a remote server. Later the server can (publicly) compute m, p= $f(m_1,...,m_l)$ together with a succinct tag σ certifying that m is the correct output of the computation f. A nice feature of homomorphic authenticators is that the validity of this tag can be verified without having to know the original dataset. But, it only integer computation and it does not support computations over real numbers.

2.1 Notations

m= Dataset (Collection of files)
σ = Authentication tag contain information about data owner and data user
p = Process
f = Computation function

3. RELATED WORK

1] Title: Efficient and Secure Storage for Outsourced Data: A Survey
Author: Jianfeng Wang, Xiaofeng Chen
Year: 2016
Description: Despite the tremendous benefits, the outsourcing paradigm brings some new security challenges. On the one hand, the cloud server may be not fully trusted, and face both internal and external security threats, such as software/hardware failures, compromised employees, hacker. A query on data stored on a cloud server may return an invalid search result. What's more, the cloud server may be ''semi-honest-but-curious'' and intentionally execute partial search operations in order to save its computation and communication overhead. Thus, one significant security challenge is how to achieve the verifiability of search results for data stored in the cloud. It means that the client should efficiently check the validation for the results returned by the cloud server. Specifically, the following two security requirements should be meet: (1) correctness: the result is the original data and has not been modified; (2) completeness: the result includes all the matched data satisfying the client's search request. On the other hand, with the rapid popularity of cloud computing, an increasing amount of data is being outsourced to the cloud in a exponential growth manner. Inevitably, this leads to a cost explosion of data storage. This concerns not only the cost of the hardware and software necessary for storing data, but also the rapidly growing energy consumption in storage systems. As a promising solution, data deduplication has attracted increasing attention from both academic and industrial community. Deduplication can eliminate redundant data by storing one single copy for duplicate data.

2] Title: Privacy- preserving keyword-based Semantic Search over encrypted cloud data.
Author: Xingming Sun, Yanling Zhu, Zhihua Xia and Lihong Chen
Year: 2014
Description: Here design a practical encrypted search solution that support semantic search based on semantic relatedness. Semantic search reinforce the system usability by returning the exactly matched files and the files including the terms semantically similar to the query keyword. The co-occurrence of terms is used as the metric to evaluate the semantic distance between terms in semantic relationship library (SRL). In this proposed scheme, exploit the architecture of two clouds, namely private cloud and public cloud. The private cloud performs the security- critical operations, while the public cloud performs the performance- critical operations. With the encrypted metadata set provided by the data owner, private cloud constructs the SRL and index. The SRL, which records the semantic similarity values of keywords, is store in private cloud for query extension. But the encrypted index is uploaded to the public cloud for efficient search. Thus the search operation is divided into two steps. The first step expands the query keyword upon SRL stored in private cloud. The second step uses the extended query keywords set to retrieve the index on public cloud.

3] Title: A survey on searching techniques over encrypted data
Author: Ms. Archana D. Narudkar, Mrs. Aparna A. Junnarkar
Year: 2015
Description: Due to a revolutionary change in the field of industries over past decade, there has been increase in demand of outsourcing of data over a wide range of network. In order to manipulate this huge amount of data in cost effective manner enterprise has adapted a prevalent technology called cloud computing that remove the burden of data management. In this data driven environment enterprise tend to store their data onto cloud that compromise of valuable asset of customer data like emails, personal health data etc. Cloud computing is turning out to be most essential paradigm in the development of information technology which offer flexible access, ubiquitous, on demand access and capital expenditure saving. Despite its technical advantage in business, enterprise should always keep concern of its privacy from the prying eyes over a network. Privacy preserving is one of the major hurdles in cloud for user, especially when the user data that reside in local storage is outsourced and computed onto cloud. The sensitive data that a cloud service provider is holding could be secure by firewalls ,intrusion detection system also CSP has full control over the infrastructure of cloud including lower level of system stack and system hardware. Although mitigate concern are taken still privacy breaches is likely to occur in the paradigm. In few cases the service provider is not fully trusted, but still need the service. Therefore, some method should be empowered to protect the user data and user queries from unauthorized person in the cloud environment. Thus, before sending data onto the cloud, data must be encrypted to protect from data privacy and unsolicited access.

4] Title: Survey on Data Storage and Retrieval Techniques over Encrypted Cloud Data
Author: Raghavendra S, Chitra S Reddy, Geeta C M, RajkumarBuyya, Venugopal K R, S SIyengar, L M Patnaik
Year: 2016
Description: As a kind of emerging business computational prototype, Cloud Computing distributes computation task on the resource pool which consists of a large number of computers and accordingly the application systems gain the computation working strength, the storage space and software service according to its demand. The working of cloud computing can be viewed by two distinctive features One is the cloud infrastructure which is the building block for the upper layer cloud application. The other is the cloud application. Cloud computing has achieved two important goals for the distributed computing by the means of three technical methods. High Scalability the cloud infrastructure can be expanded to very large scale even to thousands of servers and high Availability so that the services are available even when quite a number of servers fail. The present-day achievements in data, mobile, wireless and Internet technologies cannot be magnified. And hence Cloud computing is an emerging commercial model that promises to eliminate the need for maintaining expensive computing facilities by companies and institutes alike. Cloud computing technology makes it possible develop and host an application design for the internet where information technology (IT) related facilities are provided "as a service"; allowing clients to access technology-enabled services more economically and flexibly on a pay-as-you-use basis. Cloud Computing applications are cloud based services also known as Software as a Service (SaaS). These applications can do everything from keeping track of notes to accounting. Cloud applications give operatives access to their information from anywhere around the globe and must require an Internet connection. This ensures team work, allowing collaborated working as multiple people can view and edit the same information at once. Cloud applications also allow enterprises to push new developments to all users at once, ensuring all round benefit at the same time.

5] Title: Privacy-Preserving Outsourcing of Data Mining
Author: Anna Monreale, Wendy Hui Wang
Year: 2016

Description: In recent years, there has been considerable interest in the so-called data mining-as-service (DMaaS) paradigm for enabling organizations with limited computational resources and/or data mining expertise to outsource their data mining needs to a third party service provider. With the aid of the DMaaS paradigm, consumers no lon ger need to invest heavily or encounter difficulties in building and maintaining complex IT infrastructure. This is extremely useful for the users, for instance, some small/median-size enterprises that have limited resources but computationally expensive data mining needs for their data. Although DMaaS offers a cost-effective solution for the data owner, sharing data with a third-party service provider and allowing it to take custody of personal data raises questions about privacy protection. It shows an example scenario that a government agency presents a search warrant to the cloud service provider that has possession of individual data. As the service provider is presumably less likely to contest the order, it will release the data without informing the data owners. This situation gets even worse as some service providers secretly sell their hosted data to make profit. As large amounts of outsourced data in the DMaaS paradigm contain personal information that are in non-aggregate format, sharing the data with third-party DMaaS service providers without careful consideration will raise great threat to data privacy.

4. PROBLEM STATEMENT:

Although cloud computing offers flexible computation and storage resources, it poses challenges on verifiability of computations and data privacy. In this work focus is on investigation verifiability for privacy-preserving multi-keyword search over outsourced documents by using homomorphic MAC technique. As the cloud server may return incorrect results due to system faults or incentive to reduce computation cost, it is critical to offer verifiability of search results and privacy protection for outsourced data at the same time. To fulfill these requirements, in this system design a Verifiable Privacy-preserving keyword Search scheme, called VPSearch, by integrating an adapted homomorphic MAC technique with a privacy-preserving multi-keyword search scheme.

5. PROPOSED SYSTEM:

In proposed system use multi-keyword rank by using this if data user search any keyword in outsource data the files which contain related keywords also display, the files which contain most of the words will also be rank forward. e.g. if you search for keyword 'Protocol' then system also return the files which contain 'internet', 'network', 'authentication'. Heredesign an efficient, verifiable and privacy-preserving multi-keyword ranked searchable encryption (MRSE) scheme for outsourced cloud data under the partially honest cloud server model. It is realized by

integrating an adapted homomorphic MAC technique with a privacy preserving multi-keyword search scheme. The proposed scheme is very efficient as it relies on only one-way function for security. In this system provide detailed analysis on security, privacy, verifiability and efficiency of VPSearch. Specifically, the underlying homomorphic MAC scheme used in VPSearch can be proved to be secure. Here implement VPSearch using java for implementation and evaluate its performance over three UCI (Unique Client Identification) datasets. VPSearch is very efficient on authentication tag generation and keyword search operations.

6. ARCHITECTURE:

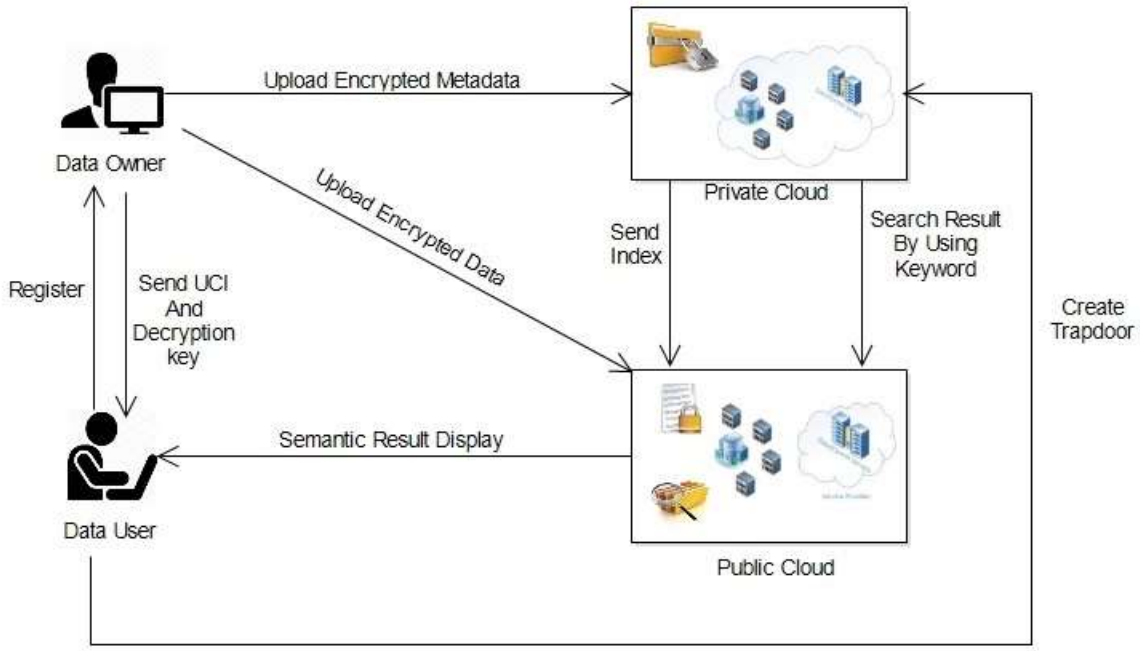

Fig -1 Architecture

Data Owner: Data owner is who creates his data and upload on cloud.
Data User: Data user is who use data owner data by using decryption key.
Private Cloud: also known as an internal or enterprise cloud, resides on company's intranet or hosted data center where all of your data is protected behind a firewall.
Public Cloud: Your data is stored in the provider's data center and the provider is responsible for the management and maintenance of the data center.
Operations Details:
1] Data owner encrypt data and upload it on public cloud. Then he creates metadata for encrypted data.
2] He encrypts created metadata and uploads it on private cloud.
3] Encrypted metadata index uploaded on public cloud.
4] When data user wants to search data owner data, first he needs to create trapdoor.
5] After successfully creation of trapdoor data user search data over public cloud using Keywords.
6] This system use homomorphic MAC technique and multi keyword search technique to show semantic result.
7] Data user registers him on data owner side and request for data decryption key.
8] By using decryption key data user decrypt required data and use it.

7. ALGORITHM:

A] RC6:

RC6Algorithm.
– RC6 is a symmetric key block cipher derived from RC5.
– Block size of 128 bits. Flexibility of key size.

– No key separation. Operators involved are simple in function favorably.
– High speed with minimal code memory.
– Provides a solid well-tuned margin for security against well-known differential & linear attacks.
– Max potential for parallelism when multiple streams are processed.
RC6 algorithm basic operations:
– a + b: integer addition modulo 2w.
– a - b: integer subtraction modulo 2w.
– A ^b: bitwise exclusive-or of w-bit words.
– An x b: integer multiplication modulo 2w.
– a¡¡¡b: rotate the w-bit word a to the left by the amount given by the least significant law bits of b.

B] AES
AES Algorithm
• AES is a symmetric block cipher that it uses the same key for both encryption and decryption.
• The AES standard states that the algorithm can only accept a block size of 128 bit.
• The entire data block is processed in parallel during each round using substitutions and permutations.
• Single 128 bit block in decryption and encryption use as input and is known as the in matrix.
Inner Workings of a Round: The algorithm begins with an Add round key stage followed by 9 rounds of four stages and a tenth round of three stages. For both encryption and decryption this applies with the exception that each step of a round the decryption algorithm is the opposite of its counterpart in the encryption algorithm. The four steps are as follows:
1. Substitute bytes.
2. Shift Rows.
3. Mix Columns.
4. Add Round Key.
The tenth round simply leaves out the Mix Columns stage. The first nine rounds of the decryption algorithm consist of the following:
1. Inverse Shift rows.
2. Inverse Substitute bytes.
3. Inverse Add Round Key.
4. Inverse Mix Columns.

C] Symmetric Encryption algorithm:
Symmetric algorithms convert plain-text data into an unreadable cipher text using a single key or password; they decrypt the cipher text using the same key.

8. METHODOLOGY IMPLEMENTATION

1] Client first encrypts the plaintext (Normal Information) then create index (i.e. metadata) and encrypt it.
2] Encrypted index is authenticated with homomorphic MAC technique. This produces authentication tags for the encrypted index.
a] The holder of a dataset{m1,..., m} uses her secret key sk to produce corresponding tags (σ1,...,σl) and stores the authenticated dataset on a remote server.
b] Later the server can (publicly) compute m p= f (m_1,..., ml) together with a succinct tag σ certifying that m is the correct output of the computation f.
c] A nice feature of homomorphic authenticators is that the validity of this tag can be verified without having to know the original dataset.
3] Next, the index and authentication tags are uploaded to the cloud. Then the client can generate a search trapdoor, and uses our homomorphic MAC technique to authenticate the trapdoor.
4] With the authenticated trapdoor, the cloud server can homomorphically execute the search function over the authentication tags to derive the result with a proof, which can certify the search result.

9. CONCLUSION

In this study we have proposed the first verifiable privacy preserving multi-keyword search scheme for cloud computing. Our scheme is implemented by applying an efficient homomorphic MAC scheme on a privacy-preserving multi keyword scheme, and we have made necessary modifications to the homomorphic MAC scheme so that it supports privacy preserving multi-keyword search. We also provide a verification technique called random

challenge with ordering for top*k* search results. We have analyzed security of our scheme and showed that it fulfills the requirement of verifiability, efficiency and data/keyword privacy.

REFERENCES

[1]. Efficient and Secure Storage for Outsourced Data: A Survey by Jianfeng Wang1 Xiaofeng Chen1 in 2016.
[2]. Security and privacy of sensitive data in cloud computing: a survey of recent developments by Ali Gholami and Erwin Laure in 2015.
[3]. An efficient and secure privacy-preserving approach for outsource data of resource constrained mobile device in cloud computing. By Syam Kumar Pasupuleti, Subramanian Ramalingam, RajkumarBuyya in 2016.
[4]. Survey on Data Storage and Retrieval Techniques over Encrypted Cloud Data by Raghavendra S, Chitra S Reddy, Geeta C M, RajkumarBuyya, Venugopal K R, S SIyengar, L M Patnaik in 2016.
[5]. Privacy-Preserving Outsourcing of Data Mining by Anna Monreale, Wendy Hui Wang in 2016
[6]. Homomorphic Signatures and Message Authentication Codes by Dario Catalano in 2014
[7]. Non-Interactive Verifiable Computing: Outsourcing Computation to Untrusted Workers
By Rosario Gennaro, Craig Gentry, and Bryan Parno
[8]. Practical Techniques for Searches on Encrypted Data By Dawn Xiaodong Song David Wagner Adrian Perrig
[9]. A Survey on Searching Techniques over Encrypted Data by Ms. Archana D. Narudkar, Mrs. Aparna A. Junnarkar in 2015
[10]. Enhancement of Cloud Computing Security with Secure Data Storage using AES by Vishal R. Pancholi and Dr. Bhadresh P. Patel in 2016