Secure Search over Encrypted Data in Cloud Computing: A Survey

Sumit Devray[1] Awdhesh Kumar[2]
[1]M.Tech. Scholar [2]Assistant Professor
[1,2]R.I.T. Indore

Abstract— Cloud computing becomes more informative and service oriented technology now in these days. A number of individual and organizations are placed their data on cloud storage. But storage facilities on cloud need a secure environment thus a number of cryptographic approaches are applied on data to keep them safe. Such kind of cryptographic approach preserve data but during information retrieval from such kind of data source becomes less feasible. Thus the presented paper provides a survey on the searchable cryptographic approaches that are efficiently retrieve data from the cryptographic data storage. Additionally a new scheme is also proposed that help to secure data on cloud storage as well as provide an efficient technique to retrieve accurate user data.

Key words: Cloud Computing, Security, Privacy Preserving, Information Retrieval, User Information

I. INTRODUCTION

Cloud in a terms defined by internet based applications. User send requests for services and server serve the desired data or application as services. Here, cloud becomes an organization of network and high computational hardware resources. These resources are connected for providing a high efficient computational experience. The cloud makes it possible for access information from anywhere and anytime. Traditional computing requires same location of data storage [1]. Some key benefits to use cloud computing which is given in [2] are as:

1) Reduced Cost: Cloud technology is a pay as need, this technique help to save money.
2) Increased Storage: Organizations can store more data than on private computer systems.
3) Highly Automated: IT personnel not needed to keep software up to date as maintenance is the job of the service provider on the cloud.
4) More Mobility: Employees can access information wherever they are, rather than having to remain at their desks.
5) Allows IT to Shift Focus: No longer having to worry about constant server updates and other computing issues, government organizations will be free to concentrate on innovation.

In addition of some limitations are seen in [2], the interesting thing about cloud computing is that cloud computing to include everything [3]. But certainly shifting to cloud computing has other problems including:

1) Security: Is there a security standard?
2) Reliance on 3rd Party: Control over own data is lost in the hands of a "difficult-to-trust" provider.
3) Cost of transition: Is it feasible for me to move from the existing architecture of my data center to the architecture of the cloud?
4) Uncertainty of benefits: Are there any long term benefits?

After discussing about the cloud computing we are going to provide the details about the working domain.

II. BACKGROUND

There is a number of issues and currently on the cloud computing systems. But security and privacy is a most crucial issue. In computing technology It is clear that security issue has played the most important role in Cloud computing. Without any fear put your data, running your software at someone else's hard disk and CPU. Additionally security issues such as data loss, phishing and botnet, pose serious threats. Additionally distributed and pooled computing resources have introduced new security challenges [4]. For example, hackers are planning to use Cloud to organize botnet as Cloud [4]. Distributed model has created two issues first, shared resources on same machine invites unexpected inference between malicious and legitimate resource. And "reputation fate-sharing" will affect reputation, unfortunately.

Thus for preserving the security and privacy of data the cryptographic techniques are utilized. The art of preserving information in to cipher text using a secret key to decipher is known as cryptographic approach. Cryptanalysis or code-breaking techniques can break the ciphers. Modern cryptography is virtually unbreakable. Cryptography systems can be broadly classified into symmetric-key systems that use a single key that both the sender and recipient have, and public-key systems that use two keys, a public key known to everyone and a private key that only the recipient of messages uses [5].

A. Cryptography

There are various different cryptographic schemes are available. Based on nature Encryption algorithms can be classified in two categories [6].

1) Symmetric Encryption:

That is a classical and well-known technique. A secret key such as a number, word or just string of random letters used to transform a message. This might be simple shifting a letter by a number.

2) Asymmetric Encryption:

The secret key exchange over Internet while preventing them, falling into wrong hands. Asymmetric encryption is effective solution there are two key pair is used. Public key available to anyone who wants to send a message. A second, private key is secret. Any message that encrypted using public key can only decrypted using private key. A message encrypted by using the private key. A problem with asymmetric encryption is that it is slower than symmetric encryption.

III. RELATED STUDY

This section reports different techniques that support privacy preserving technique over cloud computing.

Cloud Computing is a vision of computing as a utility. Users can store data into cloud to enjoy on-demand high quality applications and shared pool computing resources. During data outsourcing users relieved from

burden of local data storage and maintenance. Additionally data integrity in Cloud is very challenging task. Thus, enabling public audit ability for cloud storage security is of critical importance to check the integrity of data when required. Third party auditor (TPA), following two fundamental requirements: 1) TPA should be able to efficiently audit cloud storage without demanding local copy of data, and no additional burden to user; 2) Auditing process should bring in no new vulnerabilities towards user data privacy. Cong Wang et al. [7] utilize uniquely combine public key based homomorphic authenticator with random masking. This achieves privacy-preserving public cloud auditing system, which meets all requirements. To support efficient handling technique of bilinear aggregate signature is used for multi-user. Security and performance analysis shows the provably secure and highly efficient technique.

Data owners are motivated to outsource their data management systems from local to commercial public cloud for flexibility and cost savings. Protecting data privacy over sensitive data needs to encrypt before outsource, which obsoletes utilization of plaintext during keyword based search. Thus, to enabling an encrypted cloud data for search is a complex task. It is necessary to allow keywords search. In this paper, D. Boneh et al. [8] proposed Public-Key Encryption with Keyword Search (PEKS), in which CS contains encrypted files and keyword. For searching W user creates keyword trapdoor Tw using its private key. The CS checks trapdoor Tw with existing encrypted keyword and sends encrypted file that matches it. Since sender makes file encryption and the user authentication done by server, there exists a secure channel between the server, sender and user. Majid Bakhtiari et al. [9] Proposed PEKS secure searchable encryption algorithm based on public key encryption with key word search uses public key A_{pub} for encrypting data and keywords. For decrypting information or generating trapdoor Tw, it uses two different private keys. Schema is based on bilinear map. This encryption scheme improves the security of searchable encryption from IND-CPA to IND-CCA2. Mohammad Mahdi Tajiki et al. [10] Introduce improved version of SPKS algorithm and the security of the proposed scheme is analysed. In this paper we show that due to lack of client signature in the SPKS, an attack called forging attack is applicable on it. Taking this drawback into account, we present an attack to the SPKS algorithm. Due to imperfect signature, attackers can forge the client and produce valid cipher text without user's private key. In order to make these algorithms robust to such attacks, new method called Improved SPKS (ISPKS). For partially decryption of the cipher texts the encryption scheme employs on CSP. Consequently, the client communication and computational cost in decryption will be reduced. Although the cloud server provider participates in the deciphering process, it cannot detect any information about the plaintext. *Ning Cao et al. [11]* define a privacy-preserving multi-keyword ranked search over encrypted data (MRSE). Author establishes a set of privacy requirement to secure data. They choose efficient similarity measurement technique known as "coordinate matching," to capture relevance documents for search query. First idea for MRSE based on secure inner product computation, and then two improved MRSE schemes to achieve privacy in two different threats. To improve search experience, they extend both schemes to support more search semantics. Analysis shows privacy and efficiency guarantee.

Cloud storage enables storing of data over server efficiently and offer to work with data without any issue of resources. In existing system data stored in cloud using dynamic operation which offers to user to make a copy for updating and verification of data losses. An efficient storage auditing mechanism is used to overcome limitation in data loss. C. Selva kumar et al. [12] introducing a partitioning method for data storage to avoid local copy at user side. This method ensures storage integrity, enhanced error localization and easy identification. To achieve this, data integrity checking concept is used. In nature data are dynamic; thus main aims to store data to reduce time and computational cost.

Individuals and organizations outsource their data from local to remote servers. Additionally cloud infrastructure and platform providers, such as Amazon, Google, and Microsoft, are offering accessible and user friendly data storage to cloud users. Storing data in third party's cloud causes serious concern over data privacy. Encryption schemes used to protect data confidentiality, but also limit other process related to storage. W. Sharon Inbarani et al. [13] propose a threshold proxy re-encryption scheme and integrate with decentralized and secure distributed storage. Distributed storage not only support security and data storage and retrieval, and also support user forward his data to storage for another user.

According to Emiliano Miluzzo et al. [14] Cloud service providers invest efforts to design, develop and empower cloud infrastructures. Additionally, technical development enhances device for powerful computation, storage, and communication. Is devices extends their boundaries of cloud model to form more flexible, resource-aware, and better-performing cloud? Kalyani Bangale et al. [15] present a method to secure data server by backups. The Objective of Data Collection Server is to provide Auto Response, Solutions for Data Backup and Restore. The SRHCDCS can collect data and send to a centralized repository in independent format. The central repository is a source for other vendors to use information for their specific needs. The purpose of SRHCDC Server is to help users to collect information from remote location even if network connectivity is not available.

Hand held devices such as smart phones have increasing and became powerful. Smart phones are not only delivering voice but also equipped with wide capabilities. Mobile cloud enhances their scalability and security. The primary objective of V. Malligai et al. [16] "cloud based mobile data storage system" is to create an Android app that can store all kind of mobile data in cloud and can access anywhere any time. Thus it reduces overhead of mobile additionally making our data secure and flexible.

The increasing network bandwidth and reliable yet flexible network connections make it possible users can subscribe high quality services for data and software. To maintain data securely in distributed environment P. Srinivas et al. [17] propose an effective and flexible distributed scheme with Token Generation algorithm for checking secure and dependable cloud storage. Scheme was introduced to encrypt data with user specified key to make resource healthy. They derive an algorithm which is light

weight and easy. Encrypted blocks stored in cloud and perform token checking on encrypted blocks to verify data effectively in case of any modifications. The scheme is highly efficient and resilient against attacks like Byzantine server failures. Two way verification of file blocks which results more robust and ensure that data will not be modified before reaching to clouds.

The documents are outsourced to cloud for reducing management cost and ease of access. Although encryption protecting data privacy, but search over encrypted data is a challenging task. Wenhai Sun et al. [18] present a privacy-preserving multi-keyword text search (MTS) with similarity-based ranking. The scheme builds a search index based on term frequency and a vector space model with cosine similarity measure to achieve search accuracy. To improve search efficiency, a tree-based index and multi-dimensional (MD) algorithm is used. Further search privacy is introduced to meet privacy requirements under threat models.

The cloud storage based information retrieval service is a promising technology. Dong young Koo et al. [19] propose an efficient data retrieval scheme using attribute-based encryption. The scheme is best suited for cloud storage with massive data. It provides rich access control and fast search. The scheme also guarantees security and user privacy during retrieval process.

This section discussed a different kind of techniques and algorithms that are used for privacy and security on data. In further discussion a new proposed scheme is introduced.

IV. PROPOSED WORK

Cloud Computing is next generation architecture for computing, on-demand, self-service, network access, location independent, resource pooling, resource elasticity, usage-based pricing. Cloud Computing is transforming nature of businesses using IT. Key aspect of paradigm shifting is that data is being centralized or outsourced into Cloud. Storing data in cloud provide benefits: relief of storage management, universal data access with independent of locations, and avoidance of capital expenditure on hardware and software [7].

In order to provide information retrieval function over cloud while security and privacy is the key area of concern. The concept of searchable encryption is required to introduce. In the proposed search system some additional encrypted index terms are used. The entity uses a cryptographic algorithm similar to decryption for finding the correspondence between the encrypted query and the encrypted data content. Searchable encryption able to resolve the following privacy and security issues:

1) Privacy of data: During retrieval not any user can decipher the data.
2) Privacy of the data owner: No one can get actual identity of data owner from encrypted content.
3) Privacy of the retriever: The identities of the target retrievers are not getting from the contents.

The cloud environment provides support for efficient computing and enables to provide the efficient computing and storage. In this presented work the main aim to address the following issues:

1) Data security: The data is placed on cloud is not much secured due to third party access and malicious treads can harm the security of data, therefore data security in cloud storage is primary aim of the presented system.
2) Data owner and client privacy management: Data access identity of data owner is not distinguishable using data query and information retrieval. Additionally, privacy on such data during access is also a key aim.
3) Searchable data space: The cryptographic technique transforms data to unreadable formats and during information retrieval it creates hurdles. Thus improving accuracy of data search is also a goal of the presented work.

In order to achieve the established goal the following solutions are included for solution development.

1) Authentication management: A third party auditor is introduced during authentication process. Additionally using user attributes and one time password included managing secure authentication.
2) Cryptographic data security: for securing data from untrusted users the CPABE and Blowfish based cryptographic algorithm applied for providing security.
3) Providing the search solution over the encrypted data: the keyword based search system based on term frequency is used to identify data during data retrieval processes.

The overview of the proposed secure search system can be understood through the given figure 1.

In this diagram a third party auditor is used that is remain ideal in initial phases. First a user makes a request from the server to authenticate the user. The primary server initiate the third party auditor initiate a authentication process through a web service. When the user authenticated through the user id and their password a onetime password is generated for secure session. This password is active for only a single session and then the user get access to the system. During this process the third party auditor send the user attributes to the primary server which is used to encrypt and decrypt data for storage. There are primary three basic operations are demonstrated first the upload, download and third search.

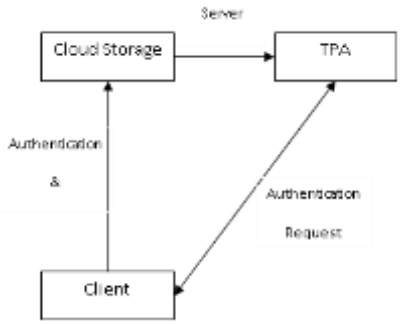

Fig. 1: server initiation

A. Upload:

During data uploading function the input text data is processed and first the stop word is removed and the tokenization is taken place. After tokenization the frequency of all data is estimated. After measuring the frequency of text data tokens the data tokens are sorted according to the

frequency and most frequent tokens are get selected as keywords for search process.

B. Download:

In this process after authentication the user simple need to select the file from a system defined list which is available for download. User simple select the file and get access the files which are owned by self.

C. Search:

In this phase the search operation is performed on the encrypted data. Thus the preserved keyword as used with the queried data sequence. In order to perform the search operation the KNN based search process is used that find the data more accurately based on single as well as the multiple keywords.

In further the encryption and decryption techniques are discussed for securing data and their privacy.

The encryption process of the proposed secure cloud storage is given using figure 2, in this diagram the after authentication process user select a text data from the local storage, during the upload operation first the data is pre-processed and the stop words are removed from the input data and then after the tokenization is performed. After tokenization the frequency of each token is computed. Using this frequency count the tokenized data is sorted and a fixed length size of data is prepared. Here the key generation process is taken place for encrypting the data, thus the selected keywords, OTP (one time password) and other user attributes are used with the CPABE algorithm to generate the secure key. Now the original input text file and the generated key is produced over the Blowfish algorithm for encryption.

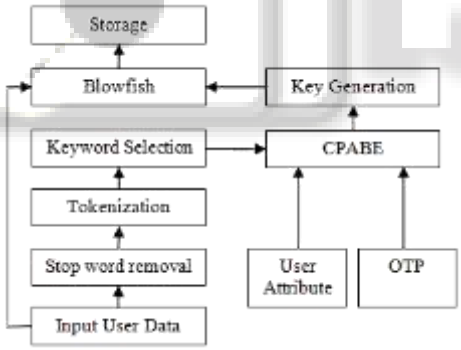

Fig. 2: data encryption

A cipher text-policy attribute based encryption scheme consists of four algorithms: Setup, Encrypt, KeyGen, and Decrypt.

Setup: The setup algorithm takes security parameter as input. It outputs the public parameters PK and a master key MK.

Encrypt (PK, M, A): Over the universe of attributes the encryption algorithm takes as input the public parameters PK, a message M, and an access structure A. The algorithm will encrypt message M and produce a cipher text CT such that only a user that possesses a set of attributes that satisfies the access structure will be able to decrypt the message.

Key Generation (MK, S): The input of key generation algorithm is master key MK and a set of attributes S that describe the key. Outputs will be a private key SK.

Decrypt (PK, CT, SK): Input of the decryption algorithm is public parameters PK, a Cipher text CT, an access policy A, and a private key SK, which is a private key for a set S of attributes. For decrypting the cipher text and getting a message M, the set S of attributes must be satisfies the access structure A.

The given section demonstrates the proposed system and their basic operations. In next section the entire work is summarized as the conclusion additionally the future extension of the work is also suggested.

V. Conclusion

Now in these days a number of efforts are made to ensure the end client for security of their data and preserving their privacy. But most of remote security based schemes are utilizing the cryptographic approach for securing data from threads and malicious users. During the cryptographic process the leave their actual format and transformed into a unreadable format. Thus the information retrieval systems are not functioning properly. Therefore in order to enhance the search relevancy of the encrypted data search a novel procedure is proposed in this work and a step procedure is also demonstrated.

In near future the given scheme is implemented using the JAVA based framework and their security and performance in terms of search relevancy is presented.

References

[1] Alexa Huth and James Cebula, “The Basics of Cloud Computing”, © 2011 Carnegie Mellon University, Produced for US-CERT.
[2] Nariman Mirzaei, Cloud Computing, Fall 2008, Community Grids Lab, Indiana University Pervasive Technology Institute.
[3] Mike Ricciuti, “Stallman: Cloud computing is 'stupidity'”, http://news.cnet.com/8301-1001_3-10054253-92.html.
[4] Y. Chen, V. Paxson, and R. Katz, "What’s New About Cloud Computing Security?” 2010.
[5] Santosh Kumar and R. H. Goudar, “Cloud Computing – Research Issues, Challenges, Architecture, Platforms and Applications: A Survey”, International Journal of Future Computer and Communication, Vol. 1, No. 4, December 2012.
[6] Rajnish Noonia, “.Net Cryptography (Encryption / Decryption)”, http://www.pixytech.com/rajnish/2013/04/ net-cryptography encryption decryption.
[7] Cong Wang, Qian Wang, and KuiRen, Wenjing Lou, “Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing”, 978-1-4244-5837-0/10/$26.00 ©2010 IEEE.
[8] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search," in Advances in Cryptology – Eurocrypt 2004, vol. 3027, pp. 506-522, Interlaken, Switzerland, 2004.
[9] Majid Bakhtiari, Majid Nateghizad and Anazida Zainal “Secure Search Over Encrypted Data in Cloud Computing” 2013 International Conference on Advanced Computer Science Applications and

Technologies 978-1-4799-2758-6/13 © 2013 IEEE DOI 10.1109/ACSAT.2013.64.

[10] Mohammad Mahdi Tajiki and Mohammad Ali Akhaee "Secure and Privacy Preserving Keyword Searching Cryptography" 978-1-4799-5383-7/14/$31.00 ©2014 IEEE.

[11] Ning Cao, Cong Wang, Ming Li, KuiRen, and Wenjing Lou, "Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data", IEEE Transactions on Parallel and Distributed Systems, Vol. 25, No. 1, JAN 2014.

[12] C. Selvakumar, G. Jeeva Rathanam, M. R. Sumalatha, "PDDS - Improving Cloud Data Storage Security Using Data Partitioning Technique", 978-1-4673-4529-3/12/$31.00 c 2012 IEEE.

[13] W. Sharon Inbarani, G. Shenbaga Moorthy, C. Kumar Charlie Paul, "An Approach for Storage Security in Cloud Computing- A Survey", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 2, Issue 1, January 2013.

[14] Emiliano Miluzzo, "I'm Cloud 2.0, and I'm Not Just a Data Center", 1089-7801/14/$31.00 © 2014 IEEE Published by the IEEE Computer Society.

[15] Kalyani Bangale, Karishma Nadhe, Nivedita Gupta, Swati Singh Parihar, Gunjan Mankar, "Smart Remote Health Care Data Collection Server", International Journal of Computer Science and Mobile Computing, Vol. 3, Issue. 2, February 2014, pg.415 – 422.

[16] Malligai, V. Venkatesa Kumar, "Cloud Based Mobile Data Storage Application System", International Journal of Advanced Research in Computer Science & Technology (IJARCST 2014) © 2014, IJARCST All Rights Reserved 126 Vol. 2 Issue Special 1 Jan-March 2014.

[17] P. Srinivas, K. Rajesh Kumar, "Secure Data transfer in Cloud Storage Systems using Dynamic Tokens", International Journal of Research in Computer and Communication technology, IJRCT, ISN 278-5841, Vol 2, Issue 1, January ,2013.

[18] Wenhai Sun, Bing Wang, Ning Cao, Ming Li, Wenjing Lou, Y. Thomas Hou, Hui Li, "Privacy-preserving Multi-keyword Text Search in the Cloud Supporting Similarity-based Ranking", ASIA CCS'13, May 8–10, 2013, Hangzhou, China. Copyright 2013 ACM 978-1-4503-1767-2/13/05.

[19] Dongyoung Koo, Junbeom Hur, Hyunsoo Yoon, "Secure and efficient data retrieval over encrypted data using attribute-based encryption in cloud storage", Computers & Electrical Engineering, Volume 39, Issue 1, January 2013, Pages 34–46.

