IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 5 Issue 10, October 2018
ISSN (Online) 2348 – 7968
www.ijiset.com

# A Survey On: Effective Content Oriented Regular Language Search for Cloud Storage

**Mrs. Vani B**
Associate Professor, Dept. of CSE, Sambhram Institute of Technology, Bangalore,
**Aman Kumar Singh**, **Anand Verma, Arun J, Kishan Kumar Mishra,**
Students, Dept.of CSE, Sambhram Institute Of Technology, Bangalore.

**Abstract -** Cloud computing enables flexible data management and ubiquitous access to data. Encryption that can be searched could simultaneously provide privacy and privacy-preserving privacy features, which is an important tool for secure storage. In this article, we provide a cloud-efficient, end-to-end encryption system that is secure against the offline keyword guessing attack (KGA). One of the strengths of the proposal compared to other existing systems is that it supports a scheme of encryption and content , or more formally a query based, file searching mechanism for textual data on a deterministic finite state machine. In addition, the concrete scheme is efficient and formally secure in the standard model.

## 1. INTRODUCTION

**C**loud storage [1] is an emerging storage model that offers cloud computing users a scalable, flexible, and cost-effective service. It also saves the user a lot of money on expensive storage devices [2]. Cloud brings convenience to customers. At the same time, there are many security and privacy issues [3][4] as data is physically stored on multiple cloud service provider servers, so customers can not fully manage their data. Customers prefer to use encryption technology to protect data privacy, which also poses another problem: how to recover data on the large volume of encrypted text. No customer can tolerate the tremendous transmission effort and waiting time for the result of data recovery. Searchable encryption technology [5], [6], [7] not only provides data encryption protection, but also enables effective search without compromising privacy. The data user generates a content token that he wants to search with his private key. As cloud computing is a difficult competitive sector, it is crucial to provide a good user experience. There is an urgent need to develop new searchable encryption schemes with meaningful search patterns for cloud storage.

### 1.1 Related Work

#### 1.1.1 Cloud Security

Although cloud computing is seen as a promising mode of service for the next-generation network, security and privacy issues are the biggest obstacle to widespread adoption in the real world. Chang et al [8]. examined the complex security of cloud computing, including firewall, access control, identity management, intrusion prevention, and converged encryption. Zheng et al [9]. proposed a mobile architecture to provide secure access from remote multimedia services. The provable data possession (PDP) is a probabilistic method of verifying cloud computing to prove the integrity of the user's data without downloading all the information.

#### 1.1.2 DFA

A multi-state finite state machine and its "control" passes from one state to another in response to external "inputs". The term "deterministic" in "deterministic finite state machines (DFAs)" refers to the fact that at each input there is a single state in which the automata can pass from their current state [10]. In 2013, Parga et al [11] showed that an automation operation could be implemented on a DFA with a polynomial time complexity, which gave rise to a two-round polynomial

minimization algorithm. Later, Sarkar et al [12] applied the EDA to online learning to enable adaptive learning. Different DFAs are designed for different chapters in online courses.

### 1.1.3 Searchable Encryption

Searchable encryption is a new technique that can protect privacy and allow search for keywords via encrypted documents. This is attracting increasing attention as the concept of Song et al [13]. Wang et al [14] used the statistical measurement approach and the one-to-many order-preserving mapping technology to create a search-ranking system on encrypted files. Liu et al [15] reduces query effort and classifies queries into multiple ranks. Xu et al [16]. Public key cryptography and fuzzy keyword search combined to design a public key encryption structure with fuzzy keyword search. However, no concrete scheme is proposed in [16] Li et al [17]. Introduce relevance scores and preference factors to enable accurate keyword searches and use classified sub-dictionaries to improve your effectiveness.

## 1.2 Motivation and Our Contributions

### 1.2.1 Motivation

The research projects mentioned in section 1.1.3 are traditional searchable encryption schemes that cannot support normal voice search. To our knowledge, Liang's schema [7] is the only search that performs a regular linguistic search in a searchable cryptographic architecture. However, a careful examination of Liang's schema [7] shows that there are some limitations that will be analyzed below.

**Small Universe Construction**: It must predefine the size of the symbol set during system configuration. The size of the public key increases when the icon set occupies a large memory space when the predefined icon set is large. To add a new icon to the system, the KGC needs to rebuild the entire system**.**

**Low Efficiency**: In ciphering and search token generation algorithms, the data user must perform many exponentiation calculations that consume energy, resulting in a lot of computational effort. In addition, the cost of the transfer is very high.

**Vulnerable to KGA:** Liang's scheme [7] cannot withstand an offline keyword (KGA) guessing attack [18], [19]. It should be noted that this is an open problem. In KGA, the attacker uses the low entropy property of the keyword to select the keyword in a much smaller space. Then, the attacker tests the guessing of the keyword guessed offline. Such an attack significantly affects the security of the searchable encryption schemes [18], [19].

### 1.2.2 Our Contributions

In this article, we design a secure data storage system that supports normal voice search. This is the protection of privacy in a standard model based on bilinear hardness Diffie-Hellman. One of the strengths of this suggestion is the normal voice search enabled, which offers a much more flexible search pattern compared to other encryption systems available for search. In our system, the encryption algorithm uses as input a public key and a standard language describing a string of arbitrary length. Then, the generated ciphertext is sent to the cloud server. In the data extraction phase, the user defines a deterministic finite state machine and generates a search token from the DFA using its secret key. The DFA defines a series of transitions, an initial state and a state of acceptance. If and only if the standard language embedded in the encrypted text is accepted by the DFA of the search token, the file is considered a match file. This test process is performed by the cloud server without knowing the plain language of the current language and DFA. In comparison with the Liang system [7], the improvements and contributions of the proposed system are listed below.

**Large Universe Construction:** The characteristic of the grand universe makes it possible to accommodate a flexible number of symbols in the system, which considerably improves the practicality of the scheme. A notable advantage is that the number of symbols is not polynomial limited, which cannot be achieved with [7]. In addition, the storage space on the wireless terminal of the user with reduced storage space is reduced. Third, the system can be easily expanded if necessary.

**High Efficiency**: The proposed regular speech search system is effective. Our scheme is built into a coupling group with a symmetrical first

order more efficient than the groups coupled with an asymmetric first order. In addition, the encryption and token generation algorithms are effective. Overhead transmission costs in the system are much lower than in [7]. The effectiveness of this system and other existing systems [5], [6], [7] is completely compared and evaluated.

## 2. PRELIMNARIES

### 2.1 Bilinear Group

Consider $G_p$ be an algorithm that on input the security parameter λ. It outputs the parameters of a prime order bilinear map as (p, g, G, $G_T$, e), where G and $G_T$ are multiplicative cyclic groups of prime order p and g is a random generator of G. The mapping e: GxG -> $G_T$ is a bilinear map. The bilinear map e has three properties: (1) bilinearity (2) non-degeneracy (3) computability.

### 2.2 Hardness Assumptions

**Assumption 1 (DBDH: Decision Bilinear Diffie - Hellman assumption)**.

Consider G be a bilinear group of prime order p and g be a generator of G. Let a, b, s € $Z_p$ be chosen at random. If an adversary A is given y = (g, $g^a$, $g^b$, $g^s$), it is hard for the attacker A to distinguish e (g, g)$^{abs}$ € $G_T$ from an element Z that is randomly chosen from $G_T$.

### 2.3 Overview of DFA

The term “deterministic finite automata (DFA)” is a terminology from the theory of computation. It accepts or rejects finite strings of symbols and executes a unique operation for each input string. A DFA M is a 5-tuple (Q, ∑, δ, $q_0$, F).

## 3. SYSTEM AND SECURITY MODEL

### 3.1 System Architecture

In this system we focus mainly on the secure data recovery function. The system consists of four entities: the Key Generation Center (KGC), the data owner, a cloud server, and the user.

- **KGC** is an abbreviation for a key generation center to which all units of the system have full confidence. KGC is responsible for generating public parameters for the entire system. In the meantime, KGC will create a public / private key pair for each legal user of the system. The private key is sent to the user via a secret channel. User public keys are published and maintained by KGC using a secure management mechanism (such as Public Key Infrastructure (PKI)).

- **The data owner** uses the cloud storage service to store sensitive personal information. The data owner uses a default language to describe the file and encrypts the default language and file, which are then offloaded to the cloud.

- **Cloud Server** provides a cloud storage service to users. Digital data is typically stored in logical pools and multiple physical servers. The cloud server is responsible for keeping the data ubiquitous and accessible to authorized users. The cloud server handles amazing data processing and computing capabilities. The cloud server responds to the data user's query and searches for the correspondence documents.

- **The data user** prompts the cloud server to perform the test calculations for the encrypted data. The data user generates a search token using his private key, which is sent to the cloud server to issue a request. The data user may use a mobile device to generate the search token so that the associated algorithms are effective and have a low transmission system time.

## 4. PROPOSED SCHEME

### 4.1 Concrete Construction

We now present our construction of efficient regular language searchable encryption system over encrypted cloud data. The Algorithms used

for the respective processes involved in the phase of development of this proposed scheme are listed here:-

- **System Setup**
- **Cloud Server Registration**
- **User Enrollment**
- **Encryption**
- **Token Generation**
- **Test**

### 4.2 An Industrial Storage Example

We give an example of industrial cloud storage to show how the proposed system works. An industrial company plans to research and develop a new product. The design process involves the following steps:

1) **Production Plan:** First, the company must carefully plan the function, structure and materials of the product, then the design information is sent to the audit department**.**
2) **Audit:** The audit department is responsible for verifying the feasibility and cost of making the system. If this is not the case, the audit service does not match the project. Otherwise, the plan is sent to the manufacturing department.

3) **Manufacture:** Upon receiving the draft proposal for the new product, the manufacturing department will create a sample according to the plan. Then the sample is sent to the test department. If the production fails, it is returned to the design department**.**

4) **Experimental Test:** The testing department checks the physical properties, function and practical use of the sample product. If the sample passes all experimental tests, it is ready for large scale production. Otherwise, the product plan would be sent back to the design department**.**

5) **Product**: After a series of research and development work, the new product can be launched on the market.

## 5. SECURITY ANALYSIS

**Theorem 1**. The proposed scheme is privacy-preserving under the DBDH assumption.

**Proof**: Assume that A is an adversary that can break the proposed scheme, then we can build an algorithm to solve the DBDH problem. Firstly, the challenger C receives the tuple (g, $g^a$, $g^b$, $g^s$, T) from the DBDH assumption.

**Theorem 2.** The proposed scheme is IND-KGA under the DBDH assumption.

**Proof**: Assume that A is an adversary that can break the proposed scheme, then we can build an algorithm C to solve the DBDH problem. Firstly, the challenger C receives the tuple (g, $g^a$, $g^b$, $g^s$, T) from the DBDH assumption.

## 6.PERFORMANCE ANALYSIS

In this section, we compare the proposed scheme with other searchable encryption schemes that supports subset keyword search [5], single keyword search [6] and regular language search [7] to evaluate the performance.

### 6.1 Comparison

Boneh and Waters [5] provides a public key system that enables keyword search, subset queries and comparison queries in a conjunctive manner. Zheng et al [6] proposed a verifiable keyword-based search function based on attributes. Liang et al. [7] introduced a standard privacy system based on Water's standard speech encryption system. This proposal is compared to these schemes in terms of functionality, overhead and overhead.

#### 6.1.1 Feature Comparison:

- **Search Function:** The present scheme [5] could support functional search patterns. The scheme [7] of our proposal allows a regular query of the language on the

encrypted indexes, which is very useful in different applications. However, the scheme [6] only allows searching with a single keyword.

- **Universe:** The universe means that the system must predefine characters during construction. This is an important factor that affects the scalability of the system. The "small universe" indicates that all supported symbols or attributes must be pre-determined. The scheme of the "Great Universe" must not predefine these symbols. It is obvious that the "Great Universe" function is better suited for a large system that can constantly add new characters. Although [7] regular linguistic research can be carried out, it is a small nonconforming construction. On the contrary, our scheme is a great universe design.

- **Standard Model:** The security test security model can be categorized as a default or random Oracle template. The proven schema based on the standard model is more secure than the random oracle model. Zheng's system [6] is based on a random oracle pattern.

### 6.2 Simulation

The PBC (Pairing Based Cryptography) library [20] is used to test the performance of schemas [5], [6], [7] and our schema. The experiments are performed on a personal laptop with the following settings: Intel Core ™ i3-2120 3.3GHz processor, 4 GB of RAM and Windows 7 64-bit operating system. We choose the elliptic curve of type A for the evaluation of the power whose expression E is: $y^2 = x^3 + x$.

### 6.3 Analysis

Comparisons and simulations show that the proposed system has a much smaller public parameter size, key text size, and token size. In addition, the encryption time and token generation time are much shorter. The design principle of this system is to reduce the transfer, storage and computational burden of data owners and users. Because the cloud server has powerful computing resources, it is acceptable to transfer some of the workload of users to the cloud server. To achieve this principle, we use the construction of a large universe during the system configuration phase so that the public parameter is short in order to save the user's memory space. In addition, the system is flexibly expandable when new features need to be added to the system. In the encryption algorithm executed by the data owner, the element is constructed**.**

## 7. CONCLUSION

In this paper, we introduce a searchable encryption scheme into the large universe to protect the security of the cloud storage system, which performs normal content oriented regular language encryption and DFA search on the cloud server. We also present a concrete design with algorithms for light encoding and token generation. The proposed scheme preserves the confidentiality and is indistinguishable from KGA, which is proven in the standard model.

## REFERENCES

1) Erl T, Cope R, Naserpour A. Cloud computing design patterns[M]. Prentice Hall Press, 2015.

2) Sookhak M, Gani A, Khan M K, et al. Dynamic remote data auditing for securing big data storage in cloud computing[J]. Information Sciences, 2017, 380: 101-116.

3) Zhang Q, Yang L T, Chen Z, Li P. Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning[J]. IEEE Internet of Things Journal, 2017, DOI: 10.1109/JIOT.2017.2732735.

4) Zhang Q, Yang L T, Chen Z, Li P. PPHOPCM: Privacy-preserving High-order Possibilistic c-Means Algorithm for Big Data Clustering with Cloud Computing[J]. IEEE Transactions on Big Data, 2017, DOI: 10.1109/TBDATA.2017.2701816.

5) Boneh D, Waters B. Conjunctive, subset, and range queries on encrypted data[C]//Theory of Cryptography Conference. Springer Berlin Heidelberg, 2007: 535-554.

6) Q. Zheng, S. Xu, and G. Ateniese. VABKS: verifiable attribute-based keyword search over outsourced encrypted data. In INFOCOM, pp. 522C530. IEEE, 2014.

7) Liang K, Huang X, Guo F, et al. Privacy-Preserving and Regular Language Search Over Encrypted Cloud Data[J]. IEEE Transactions on Information Forensics and Security, 2016, 11(10): 2365-2376.

8) Chang V, Ramachandran M. Towards achieving data security with the cloud computing adoption framework [J]. IEEE Transactions on Services Computing, 2016, 9(1): 138-151.

9) Zheng X H, Chen N, Chen Z, et al. Mobile cloud based framework for remote-resident multimedia discovery and access[J]. Journal of Internet Technology, 2014, 15(6): 1043-1050.

10) Hopcroft JE, Motwani R, Ullman JD. Automata theory, languages, and computation. International Edition 24 (2006).

11) de Parga MV, Garcła P, Lpez D. A polynomial double reversal minimization algorithm for deterministic finite automata. Theoret-ical Computer Science. 2013 May 27;487:17-22.

12) Sarkar P, Kar C. Adaptive E-learning using Deterministic Finite Automata[J]. International Journal of Computer Applications, 2014, 97(21).

13) D.X. Song, D. Wagner, A. Perrig, "Practical techniques for searches on encrypted data", in: IEEE Symposium on Security and Privacy, 2000, pp. 44-55.

14) Wang C, Cao N, Ren K, et al. Enabling secure and efficient ranked keyword search over outsourced cloud data[J]. IEEE Transactions on parallel and distributed systems, 2012, 23(8): 1467-1479.

15) Liu Q, Tan C C, Wu J, et al. Towards differential query services in cost-efficient clouds[J]. IEEE Transactions on parallel and Distribut-ed Systems, 2014, 25(6): 1648-1658.

16) Xu P, Jin H, Wu Q, et al. Public-key encryption with fuzzy key-word search: A provably secure scheme under keyword guessing attack[J]. IEEE Transactions on computers, 2013, 62(11): 2266-2277.

17) Li H, Yang Y, Luan T H, et al. Enabling fine-grained multi-keyword search supporting classified sub-dictionaries over encrypted cloud data[J]. IEEE Transactions on Dependable and Secure Computing, 2016, 13(3): 312-325

18) J. W. Byun, H. S. Rhee, H.-A. Park, and D. H. Lee, Offline keyword guessing attacks on recent keyword search schemes over encrypted data, in Proc. 3rd VLDB Workshop Secure Data Manage. (SDM), vol. 4165. Seoul, Korea, Sep. 2006, pp. 75C83.

19) W. C. Yau, R. C. W. Phan, S. H. Heng, and B. M. Goi, Key-word guessing attacks on secure searchable public key encryption schemes with a designated tester, Int. J. Comput. Math., vol. 90, no. 12, pp. 2581C2587, 2013.

20) B. Lynn. The Stanford Pairing Based Crypto Library. [Online]. Available: http://crypto.stanford.edu/pbc, accessed Dec 31, 2017