Taylor & Francis
Taylor & Francis Group

Automatika

Journal for Control, Measurement, Electronics, Computing and Communications

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/taut20

Searchable encryption algorithm in computer big data processing application

Lu Ming

To cite this article: Lu Ming (2023) Searchable encryption algorithm in computer big data processing application, Automatika, 64:4, 1204-1214, DOI: 10.1080/00051144.2023.2254978

To link to this article: https://doi.org/10.1080/00051144.2023.2254978

AUTOMATIKA
2023, VOL. 64, NO. 4, 1204–1214
https://doi.org/10.1080/00051144.2023.2254978

Searchable encryption algorithm in computer big data processing application

Lu Ming

College of Mechanical Electronic and Information Engineering, Wuxi Vocational Institute of Arts & Technology, Yixing, People's Republic of China

ABSTRACT

With the continuous development of computer technology, the amount of data has increased sharply, which has promoted more and more diversified data transportation and processing methods. At the same time, computer data analysis technology can effectively process data. This is reflected in the computer big data analysis technology not only can realize data visualization analysis, but also has data prediction and data quality management. The development of cloud computing network technology can not only provide convenience points for individuals, but also provide space for enterprises to store data. The emergence of keyword search encryption algorithms solves this problem. When users use keywords to search encryption algorithms, they can search for cipher text keywords to find the files or data they want in the cloud environment. At present, it has been widely used. In addition, this article also improves the keyword search plan and the user's query plan according to the dynamic changes of keywords, and proposes a user's multi-dynamic keyword search encryption plan. Through this program, users can search for encrypted files by keywords and change them, and the changed data will be dynamically updated. In this way, the program can realize multi-user data sharing, and can realize efficient search and dynamics.

ARTICLE HISTORY
Received 3 April 2023
Accepted 29 August 2023

KEYWORDS
Big data; searchable encryption algorithm; cloud environment; information security

1. Introduction

With the continuous development of science and technology, traditional computing modes and computing algorithms can no longer meet the needs of current computers, which requires the development of new technologies and algorithms. Cloud computing is one of the new algorithms. Its emergence solves the problems of data calculation and storage. Now it has been widely used in all walks of life. Its concept has also inspired the development of other technologies. This article focuses on cloud computing. Conducted research to provide a theoretical basis for the development of science and technology. Under the current background of rapid development of science and technology, whether the efficiency and quality of data processing can be guaranteed is the key to restricting the development of science and technology, which requires the relevant data processing technology to have high-energy data processing capabilities and secure encryption technology. Therefore, in order to achieve this goal, this article focuses on a detailed study of computer to big data processing technology and cloud computing technology, hoping to find a breakthrough for the development of data processing technology and open up a path for future development. Cloud computing is currently a hot data processing technology, which can combine storage resources, software resources, computing resources, etc. through clusters. Users can connect to the cluster through the network to obtain resources and storage space in the cluster. Whether the use of cloud computing can effectively reduce the waste of resources and increase the efficiency of computing at the same time, this is also the focus of academic research. However, there are still certain problems in the development of cloud computing technology [1]. The first thing to bear is the security issue of cloud computing. If you want to achieve the healthy development of cloud services, it is necessary to solve the security problems of existing cloud computing technologies, so that enterprises and individual users can use cloud computing to solve problems with confidence [2]. Among them, the existence of cloud storage technology allows data to be used efficiently. In order to reduce resource consumption and improve business level, small businesses or individual users often like to store data in cloud space, because the cost of storing data in cloud space is much lower than other methods. However, cloud space is a third-party storage organization, and there are still certain hidden dangers in terms of security. If users directly store data in the cloud space

CONTACT Lu Ming ynydxj@126.com College of Mechanical Electronic and Information Engineering, Wuxi Vocational Institute of Arts & Technology, Yixing, 214206, People's Republic of China

without processing, once there is a loophole in the cloud storage space, the user's information security cannot be guaranteed [3]. Therefore, it is not enough to just put the data in the cloud space, and it is necessary to rely on secure encryption services in addition to the cloud services provided by the cloud service provider. However, the encrypted data does not have the original language structure and characteristics, and the user cannot retrieve it, which is not conducive to the user's use. In order to solve this problem, the encrypted file needs to be downloaded and decrypted first, so that the encrypted data is converted into plaintext again, and it is much easier to retrieve the plaintext. The ciphertext search technology has two methods: full-text search and ciphertext search. The former directly searches the full-text to obtain the target ciphertext directly, while the latter is realized by searching keywords.

2. Related work

The literature proposes that the KES scheme has the following four steps: (1) Generate a key, and the generated key needs to be kept by the user; (2) Encryption, which is mainly to encrypt the index, and upload it to the server after being encrypted locally. In the first process, the ciphertext is not leaked; (3) Generate search credentials, and users can obtain search credentials through the key and upload them to the server. However, the fact that the server obtains the search credential does not mean that it can search the information stored by the user. (4) Search. After obtaining the search credential, the user's key is also needed to perform a ciphertext search and return the qualified data. The server cannot obtain redundant information in this process. The literature is the first batch of literature to study KES. In this literature, it includes an extension to the KES program, and a multi-keyword search method is also designed [4]. The literature found that some scholars have proposed an efficient and dynamic update searchable encryption scheme with time complexity [5]. This scheme retains the advantages of the original scheme while optimizing the encryption method of the data structure, and adds data addition and deletion query tables, and operations on files [6]. The process was documented. Later, someone improved the scheme, forming a dynamic update search scheme that supports the red-black tree structure, so that the scheme can be carried out in parallel, and the efficiency can be improved by taking advantage of the advantages of multi-processors [7]. The literature proposes the first many-to-one model of KSE scheme, which can be applied to the service scenario of routers. In this scenario, the sender and recipient of the file are two different users, and the server plays a role in it [8]. The function is to filter the router information. The literature believes that KR-PEKS is constructed in accordance with the K-resilient IBE scheme, and can be expanded on this basis: on the one hand, it can support tube detection and search; on the other hand, it can also remove the safe passage [9]. Compared with BDOP-PEKS, the processing efficiency of this scheme is high, and it does not even need to use two-line row operation [10]. But it is necessary to set the security parameters, otherwise the number of queries that can be inquired cannot be controlled [11]. When setting parameters, adjust them to appropriate sizes according to actual needs [12]. If the parameters are too large or too small, it is not conducive to the operation of the program. The literature points out whether the conversion of the IND-CKA's security plan into a security plan can meet the dual requirements of computing and security [13]. The literature believes that there are many schemes based on the many-to-one model, which will not be repeated here [14,15]. These schemes can search for keywords and perform stepwise encryption calculations on data to generate ciphertext. When searching, users can generate search credentials based on keywords and ciphertexts, so that the server can perform calculations. It should be noted that this solution relies on BDH mathematical assumptions and requires the use of the double-line feature of mathematical assumptions.

3. Research on searchable encryption algorithms under the background of 3 big data

3.1. Keyword searchable encryption basics

Definition 1: There is a set $G = \{g_1,g_2,g_3 \dots\}$. If the set G pair operation $*$ satisfies all the following conditions, it is called the group $< G,* >$:

Closeness: if g_i, $g_j \in G$, there is $gk \in G$, satisfying $g_i * g_j = g_k$

Associative law: if g_i, g_j, $g_k \in G$, there is always $(g_i * g_j) * g_k = g(_i *) (g_j * g_k)$

There is identity element: there is $g_e \in G$, for any $gk \in G$, there is always $g_k * g_e = g_k * g_e = g_k$

There is an inverse element: for any $g_i \in G$, there is always $g_i * g_j = g_j * g_i = g_e$, where g_e is the identity element.

Definition 2: The number of group elements is called the order of the group, denoted as |G|

Definition 3: There is a group $< G,* >$, where G is a non-empty set, and $*$ is an operation on the set. If there is $g \in G$, for any element $gi \in G$, it can be expressed as $g_i = g^{\wedge}n$, and n is an integer, then the group $< G,* >$ is a cyclic group, and g is the generator of the group. Research based on bilinear pairing has greatly promoted the development of cryptography research.

The Lagrange interpolation theorem can be used to realize secret sharing. Based on the theorem,

KP-ABE and CP-ABE schemes can be constructed. The Lagrange interpolation theorem is as follows.

Then the polynomial can be determined as:

$$f(x) = \sum_{i=1}^{n+1} \left(f(x_i) \prod_{1 \le j \le n+1, j \ne i} \frac{x - x_j}{x_i - x_j} \right) \quad (1)$$

The access tree can flexibly express the access authority control, so some ABE schemes are implemented by using the access tree.

When checking whether the user attribute matches the authority corresponding to the access tree T, it is necessary to let R be the root node of the T tree. If x is a non-leaf node, the child node x'of x needs to be calculated.

Cryptography relies on the difficulty of breaking through mathematical problems to ensure the security of the scheme. If a cryptographic scheme passes strict proofs and can be simplified to a certain mathematical problem, then the scheme is considered safe. This section mainly introduces the definition of the mathematical problems involved in this article.

Random prediction models can be used to analyze the security of cryptographic schemes. The random prediction model is defined as follows:

It satisfies the operation H: {0,1}^*→{0,1}^*→, and has the following three properties:

Uniformity: If the input is random, the output distribution is uniform;

Determinism: If the input is the same, the output is also the same;

Validity: the output result of the polynomial.

In order to meet certainty and uniformity, the output entropy value of the random prediction model must be less than the input entropy value, and entropy theory stipulates that the output entropy value of the deterministic function should not be greater than the input entropy value. Therefore, ROM is an ideal model and cannot be truly realized. In the implementation of the algorithm, the random prediction model is instantiated by a specific one-way hash function.

The scheme based on the standard model does not rely on a random vector machine, but only relies on breaking through mathematically difficult problems to ensure safety. Therefore, the security of the standard model is higher than that of the random vector model. It can be seen that the encryption algorithm design based on the standard model is still a research hotspot.

3.2. D-ATTR-PEKS scheme design

This chapter first introduces the classic structure of the public key-based KSE scheme, and points out the shortcomings of several schemes and corresponding solutions. The existing PEKS schemes are all extended on the basis of the scheme proposed by Boneh et al. The description of the program is as follows.

(1) Keysgen(s)→(A_{pub},A_{prv})→: s is the security parameter, (A_{pub},A_{prv}) are the public key and private key respectively

According to the safety parameters, calculate

$$A_{pub} = \{g, h = g^a\}, A_{priv} = \alpha \quad (2)$$

(2) PEKS(A_{pub},W)→S: A_{pub} is the public key, W is the search keyword, and S is the ciphertext.

The algorithm is executed by the data sender, rεZ_p^* is randomly selected, and the calculation

$$S = \{A = g^r, B = H_2(e(H_1(W), h^r))\} \quad (3)$$

(3) Trapdoor(A_{priv},W)→T_w: A_{priv} is the user's private key, W is the search keyword, and T_w is the search credential.

The algorithm is executed by the data receiver and calculates

$$T_w = H_1(W)^{\alpha} \in G_1 \quad (4)$$

This solution satisfies the security of CKA, that is, it can only guarantee the security of PEKS ciphertext, and users cannot obtain any information about keywords from PEKS ciphertext. Someone pointed out that the PEKS solution must establish a secure channel, otherwise the attacker may intercept the user's query results, and even further use the data to obtain user information. Therefore, it was pointed out that SCF-PEKS cannot resist offline keyword guessing attacks. This is because the keyword ciphertext space is not as large as the keyword space, and generally speaking, the keyword selection entropy is low, so the attacker can crack the dictionary attack.

The dPEKS solution introduces the concept of a trusted server designated by the user, allowing users to choose a trusted server to upload ciphertexts. When other users need to search, the trusted server must predecrypt the search credentials uploaded by the user. If intercepted, the attacker will not be able to obtain information about the certificate. The dPEKS scheme is described as follows:

(1) Setup(()): s is a safety parameter

This algorithm is executed by an authorized institution and can generate hash functions H_1:{0,1}^*→G_1 and H_2:G_2→{0,1}^logp

(2) $Keygen_{server}$→(sk_s,pk_s): (sk_s,pk_s) is the public and private key pair of the search server

The algorithm is executed by an authorized institution, and $\alpha \varepsilon Z_p$,$Q \varepsilon G$ are randomly selected, and the calculation

$$sk_s = \alpha, pk_r = (Q, y_s) = (Q, g^a) \quad (5)$$

Among them, sk_s is the private key of the search server, which is saved by the search server, and pk_s is public as the public key of the search server.

(3) $Keygen_{server} \rightarrow (sk_s,pk_s)$: (sk_s,pk_s) is the receiver's public and private key pair

The algorithm is executed by an authorized institution, $x \in Z_p$ is randomly selected, and the calculation

$$sk_r = x, pk_r = g^x \quad (6)$$

(4) $dPEKS(pk_r,pk_s,W) \rightarrow C$: pk_r is the server public key, pk_s is the user public key, W is the encryption keyword, C is the searchable ciphertext.

The algorithm is executed by the data sender, selects a random number $r \in z_p$, and calculates

$$C = \{A, B\} = \{(pk_r)^r, H_2(e(y_s, H_1(w)^r))\} \quad (7)$$

(5) $Trapdoor(sk_r,W) \rightarrow T_W$, sk_r is the user's private key, W is the search keyword, and T_w is the generated search credential.

The algorithm is executed by the data receiver, $r^{\wedge}{}' \in Z_P$ is randomly selected, and the calculation

$$T_w = \{T_1, T_2\} = \{y_s^{r'}, H_1(w)^{1/x} g^{r'}\} \quad (8)$$

(6) $dTest(C,T) \rightarrow b$ input ciphertext C and search credentials T_w

The algorithm is executed by the search server and first calculates

$$T = (T_2)^{\alpha}/(T_1) \quad (9)$$

The attack method of this scheme is described as follows:

The attacker is interested in the keyword set $D = \{W_1,W_2,\ldots W_i\}$, calculated separately

$$\begin{cases} C_1 = \{(pk_r), H_2(e(y_s, H_1(w_1)))\} \\ C_2 = \{(pk_r), H_2(e(y_s, H_1(w_2)))\} \\ \cdots \\ C_3 = \{(pk_r), H_2(e(y_s, H_1(w_n)))\} \end{cases} \quad (10)$$

When a user initiates a search request to the server, the attacker will monitor and intercept the search credentials. After the server receives the search credentials, it will traverse all the ciphertexts uploaded by the user and calculate them one by one, including the ciphertext uploaded by the attacker maliciously.

The attacker monitors the server and intercepts the search results it sends to the user. If the file ID sent by the attacker exists in the result, the attacker will get the keyword corresponding to the search credential and know the keyword requested by the search user. The SPEKS scheme designed by Chen et al. can be used to defend against online keyword guessing attacks. Users can use session information to decrypt search results locally to ensure that attackers cannot obtain the information.

The above solutions all take the security of the KSE solution as the entry point. If the KSE solution is deployed in a cloud environment, the multi-user situation must be considered. CP-ABE is used to solve the many-to-many access problem in the cloud environment. Access control permissions can be set for ciphertext, and file access can be performed only when the user attributes comply with the access policy. The procedure is described below.

(1) $Setup(t) \rightarrow (PK,MK)$: t is the security parameter, (PK,MK) is the system public and private key pair

The algorithm is executed by an authorized institution, and generates p-order groups G_1, G_2 according to the security parameters. At the same time, generate a random vector model $H_1:\{0,1\}^{\wedge *} \rightarrow G_1$ and a one-way hash function $H_2:\{0,1\}^{\wedge *} \rightarrow z_p$. Randomly select a, b, $c \in z_p$, and the generator $g \in G$. Calculation

$$PK = \{g^a, g^b, g^c\}, MK = \{a, b, c\} \quad (11)$$

(2) keyGen (MK, S) $\rightarrow$ SK: MK is the system master key, S is the user attribute set, and SK is the user private key

The algorithm is executed by the authorized institution, $r \varepsilon Z_p$ is randomly selected, $r_j \in Z_p$ is randomly selected for each element j in the set S, and finally generated

$$SK = \left\{S, A = g^{\frac{ac-r}{b}}, \forall j \in S, A_j = g^r H_1(j)^{rj}, B_j = g^{rj}\right\} \quad (12)$$

(3) $Enc(w,T) \rightarrow Cph$: w is the search keyword, T is the access control tree, and Cph is the searchable ciphertext

The algorithm is executed by the data sender, randomly selecting r_1,r_2 ϵZ_p, calculating $W = g^{\wedge} cr1$, $W_0 = g^{\wedge}(a(r_1 + r_2))$ $g^{\wedge}(bH_2$ (W) $R_1)$,$W^{\wedge}{}' = g^{\wedge}$ (Br_2). The access control tree is defined in the same way as the CP-ABE algorithm. The root node of the policy tree T is set to r_2, and the child nodes

are assigned random polynomials. The calculation is $w_j = g^{\wedge}(qv(0))$, $D_j = H_1(j)^{\wedge}(qv(0))$. Finally got

$$Cph = \{T, W, W_0, W', j \in Attr(T), W_j, D_j\} \quad (13)$$

(4) TokenGen (SK, w) → TK: SK is the user's private key, w is the query keyword, and TK is the search credential.

The algorithm is executed by the data receiver, randomly selects $s \epsilon Z_p$, and calculates $tok_1 = (g^a g^{bH_2(W)})^S$, $tok_2 = g^{cs}$, $tok_3 = A^s = g^{(acs-rs)/b}$. Calculate $A_j{}^{\wedge}{}' = A_j{}^{\wedge}s$, $B_j{}^{\wedge}{}' = B_j{}^{\wedge}s$ for each attribute, and finally generate search credentials

$$TK = \{tok_1, tok_2, tok_3, \forall j \in S, A'_j, B'_j\} \quad (14)$$

(5) Search (TK, Cph)→b: TK is the search credential, Cph is the key word ciphertext, output $b \varepsilon 0,1 \in$

The algorithm is the same as CP-ABE, first calculate

$$E_v = \frac{e(A_j^{'}, w_{v)})}{e(B_j^{'}, D_v)} = e(g, g)^{rsqv(0)} \quad (15)$$

Use the recursive algorithm to get $e(g,g)^{rsqv(0)} = E_{root}$. Final judgment

$$e(W_0, tok_2) = e(W, tok_1)E_{root}e(tok_3, W') \quad (16)$$

If the equation is true, output 1; otherwise, output 0. However, this solution does not consider offline keyword guessing attacks.

In the above-mentioned classic construction scheme, although PEKS implements the KSE scheme, it does not consider the problem of keyword guessing attacks. In addition, the above solutions do not support many-to-many models, so they are not suitable for cloud storage environments that support data sharing. CP-ABKS supports multiple users, but it cannot resist offline keyword guessing attacks.

The cloud storage server can store the user's resources and information. Authorized institutions are responsible for establishing encryption systems, and generating and issuing public and private keys and user keys for search servers. The data owner is the owner of the data. When the data owner needs to upload data, he can use any encryption technology to encrypt the file, and then use the scheme proposed in this article to encrypt the keywords. The search server is mainly responsible for searching. And the search server can use the search credentials uploaded by the data user to search for the ciphertext. The search server will return the correct information only when the user authority meets the encryption keyword search authority control, and the search keyword matches the file keyword.

The scheme proposed in this paper has the following 5 types of roles, as shown in Figure 1.

The solution in this paper is composed of algorithms such as search server key generation, user private key generation, keyword encryption, search credential generation, keyword search, search result encryption, and search result decryption.

This section analyzes the correctness of the Test algorithm, as shown in Equation 17. Secondly, according to Equation 18, the intermediate result Y is obtained, as shown in Equation 19. Finally, according to Equation 20, the intermediate result Z is obtained, as shown in Equation 21.

$$X = e(H, T_0) = e(g, g)^{\alpha\beta st} e(H_2(W)g)^{\alpha\beta st} \quad (17)$$

$$Y = e(W_1, T_1) = e(g, g)^{\alpha\beta st + rts} \quad (18)$$

$$V = DecryptNode(CT, SK, R) = e(g, g)^{rst} \quad (19)$$

$$Z = e(W_2, T_2) = e(H_2(W), g)^{\beta kt} \quad (20)$$

Finally, the final calculation result is obtained, as shown in Equation 20, which proves that the solution is correct.

$$\left(\frac{Y}{V}\right) = \left(\frac{X}{Z}\right) = e(g, g)^{\alpha\beta st} \quad (21)$$

Table 1 mainly compares from five aspects:

Combined with the functional analysis in Table 1, the solution in this article does not require a secure channel, can effectively resist external intrusions, and is more suitable for running in a cloud environment.

The program analyzes the execution efficiency through experiments, and selects the A-type elliptic curve in JPBC. Figure 2 shows the execution efficiency of a typical algorithm.

The experimental results show that each algorithm is roughly proportional to the number of attributes, so the time spent in the solution in this paper is within an acceptable range.

3.3. MU-DSSE scheme

Improving the search efficiency of the KSE scheme based on the public key is still a hot research topic. There are two ways to improve search efficiency: use more effective mathematical operations instead of pairing operations, and someone proposed a DSSE scheme that supports effective dynamic updates. The program also introduces a delete array and its corresponding quick reference table to save deleted file information.

The establishment of an inverted index is shown in Figure 3.

The DSSE scheme builds an index model as shown in Figure 4.

The DSSE scheme is very efficient, but it does not support multiple users. The Mu-MQ solution supports

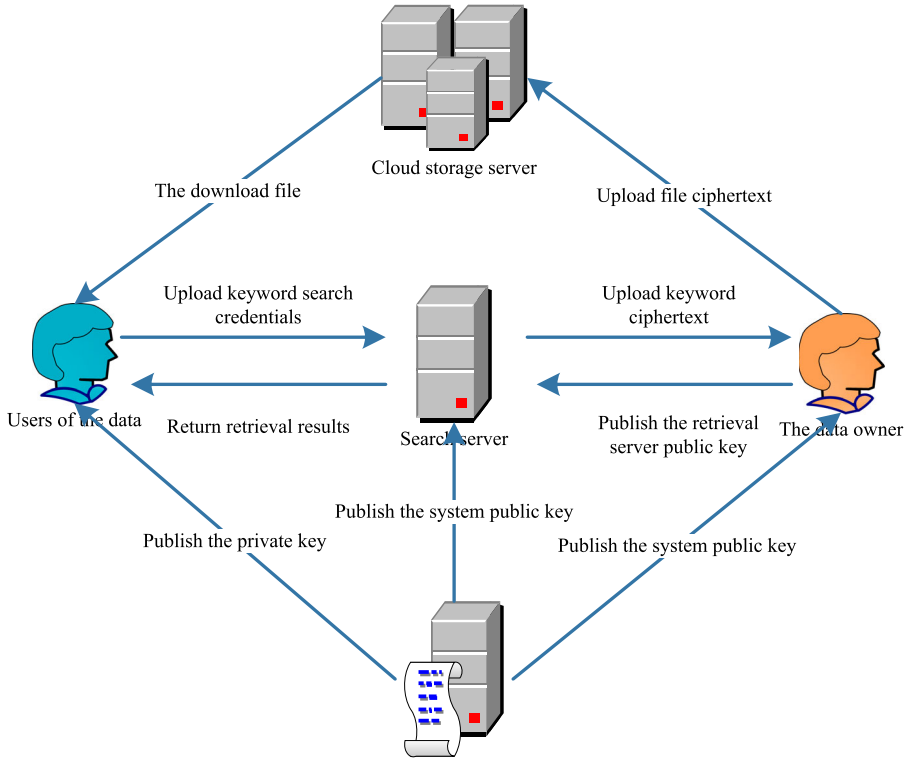

Figure 1. System model.

Table 1. Scheme function comparison.

Scheme	No need for a secure channel	Resist offline keyword guessing attacks	Resistance online key	Access control	Does not require an authority to generate search credentials
SCF-PEKS	Yes	No	Word guessing attack	No	Yes
dPEKS	Yes	Yes	No	No	Yes
SPEKS	Yes	Yes	No	No	Yes
ATT-PEKS	No	No	Yes	Yes	Yes
VABKS	No	No	No	Yes	Yes
ABEKS	No	Yes	No	Yes	No
Program in this section	Yes	Yes	No	Yes	Yes

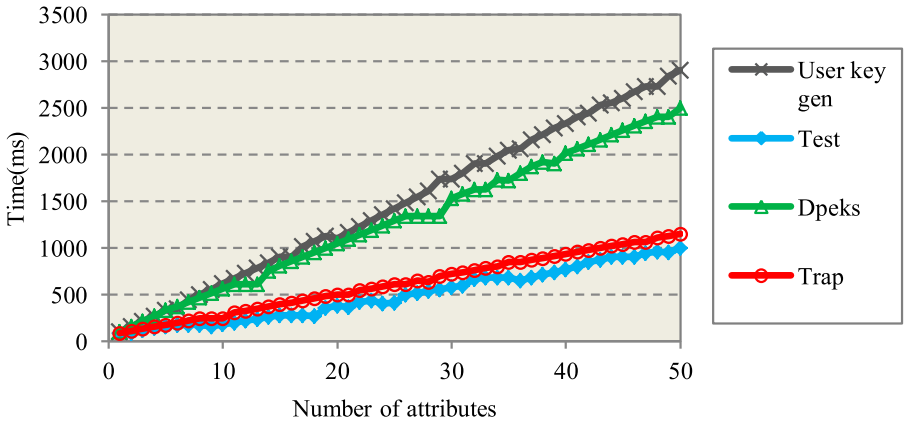

Figure 2. Experimental results.

multiple users, but it is not efficient. Therefore, the MU-DSSE scheme proposed in this section uses the idea of the Mu-PQ scheme to extend the DSSE scheme to a multi-user scheme. In this solution, a trusted group is formed for users who have access to data, and each user in this group has a key for searching. Using this keyword, you can effectively implement index search, index addition, and index deletion. The user can revoke the user or add the user to a trusted group at any time.

The scheme proposed in this section has the following three types of roles: user groups, authorized institutions, and cloud search servers, as shown in Figure 5.

The program is described as follows. Because the research focus of this article is to search and encrypt

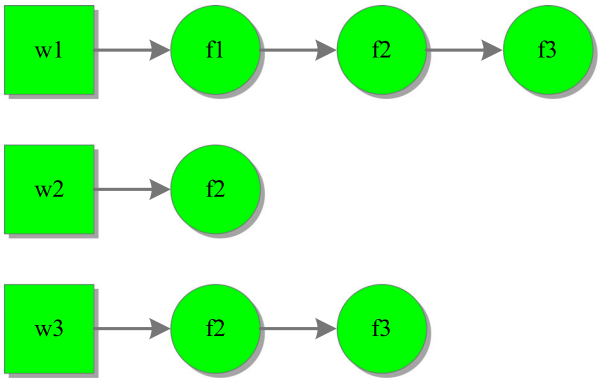

Figure 3. Inverted index.

keywords and encrypt files. In order to use any encryption method, it is not mentioned in the plan.

1. The initialization algorithm is executed by the authorized agency, and the authorized agency generates a bilinear group and combines two random vector machines

Save a, b, and c as the system key, as shown in equation 22.

$$MK = \{a, b, c\} \tag{22}$$

2. The algorithm is executed by an authorized institution, and the authorized institution distributes keys for users in the user group. Then calculate according to formula 23

$$PK_{uid} = \{AK_{uid} = g^{a/K_{uid}},$$
$$BK_{uid} = g^{b/K_{uid}}, CK_{uid} = g^{c/K_{uid}}\} \tag{23}$$

3. The algorithm is executed by the user. The user uses the private key to generate the search credential according to formula 24, and hand the search credential to the server.

$$T_w = \{uid, T = h(w)K_{uid}\} \tag{24}$$

4. The algorithm is executed by the cloud search server. The server obtains the user search credential T_W, and finds the corresponding PK_{uid} according to the uid. Calculate according to formula 25 and 26 respectively

$$F(w) + e(AK_{uid}, T_w) \tag{25}$$
$$G(w) + e(BK_{uid}, T_w) \tag{26}$$

This section compares the proposed scheme with the classic multi-user KSE scheme and explains the advantages of this scheme. Table 2 mainly compares search efficiency, system model, whether it supports dynamic update, whether it supports multi-user update, and whether it supports flexible exit.

Next, perform the efficiency through experimental analysis.

(1) O(1) algorithm efficiency

Complexity algorithm – just keep the number of pairwise operations or exponential operations, and its efficiency is shown in Table 3.

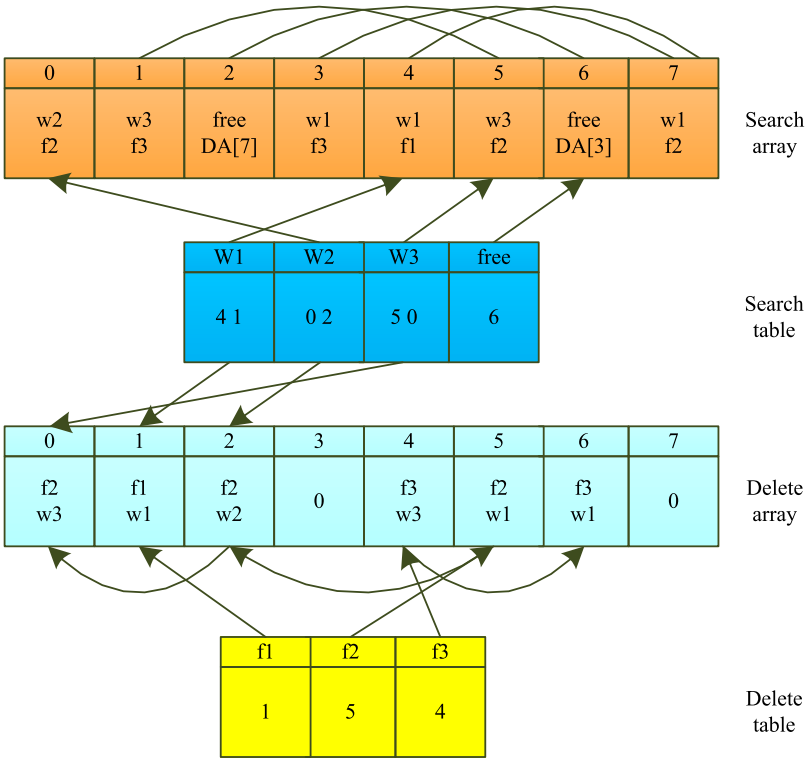

Figure 4. Index establishment.

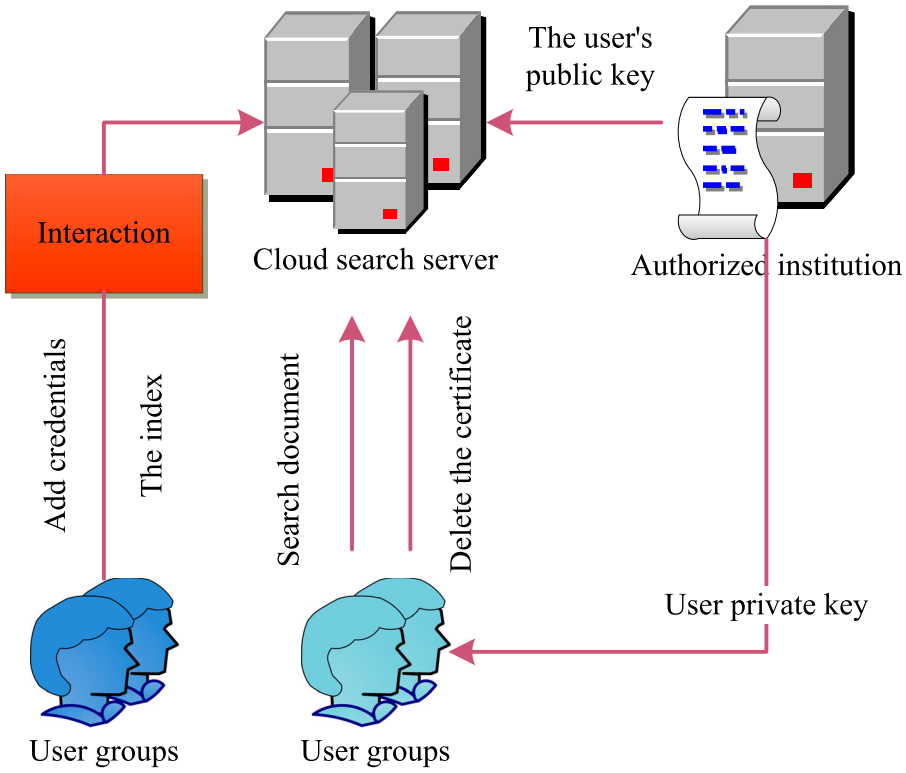

Figure 5. MU-DSSE solution system model.

Table 2. Comparison of scheme functions.

Scheme	Effectiveness	System model	Dynamic update	Multi-user update	Flexible revocation
BE SSE-1	O(1)	One to many	No	No	No
ABE SSE-1	O(1)	Many to many	No	No	No
MuPQ	O(n∗|W|)	Many to many	Yes	Yes	Yes
PEKS	O(n∗|W|)	Many to one	Yes	Yes	No
MPEKS	O(n∗|W|)	Many to many	Yes	Yes	No
Program in this section	O(1)	Many to many	Yes	Yes	Yes

Table 3. O(1) algorithm efficiency.

Algorithm	Time (ms)
Setup	522
AddUser	83
DelToken	74
Search Token	71
Delete	50

Table 4. Comparison of schemes.

	MU-DSSE	D-ATTR-PEKS
Search efficiency	o(n)	O((n-m)∗ |w| ∗|S|)
Trusted object	Group users, authorized institutions	Search server, authority
Access control	Group-based	Attribute-based
User cancellation	Stand by	Not support
Interactive execution	Need	Not needed

(2) Buildindex algorithm

When constructing the index, the algorithm needs to interact with the server, so it is divided into an interactive phase and a local execution phase. As shown in the step3 bar graph in Figure 6. The time and keywords used in the three steps of the interactive phase are roughly proportional to the total number of file IDs. The time spent in the three stages gradually decreases.

The local execution stage is the stage where the client builds and encrypts the index. The time consumed in this stage is mainly related to the length of the SearchArray and DeleteArray constructed.

(3) Add operation

In the "Add Algorithm Efficiency Test", let the server save 100,000 files, each file contains 5 keywords, of which files and keywords are randomly generated test data. The test result is shown in Figure 7. The X axis is the number of file index keywords added, and the Y axis is the algorithm execution time. According to the experimental results, the time complexity of the algorithm is O, and it can be carried out relatively quickly. According to the specific form of the algorithm, the operation of this algorithm is similar to the Delete algorithm. Because the delete algorithm needs to decrypt larger data items, it is faster than the delete algorithm.

This section compares the two schemes of MU-DSSE and D-ATTR-PEKS, and the comparison results are shown in Table 4.

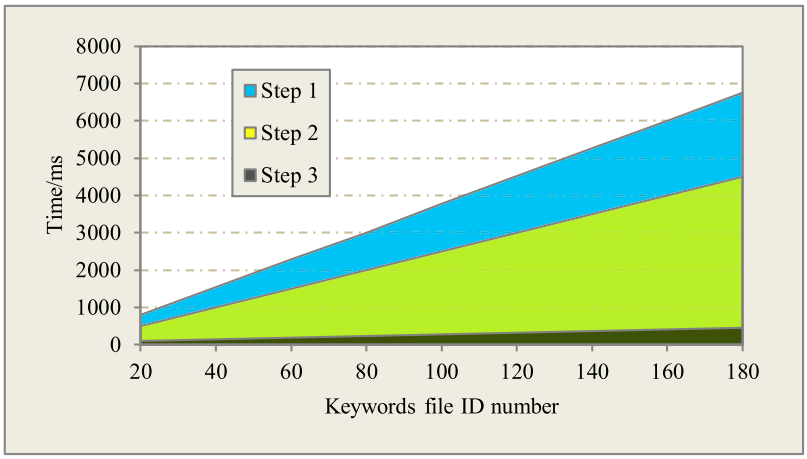

Figure 6. BuildIndex interactive phase.

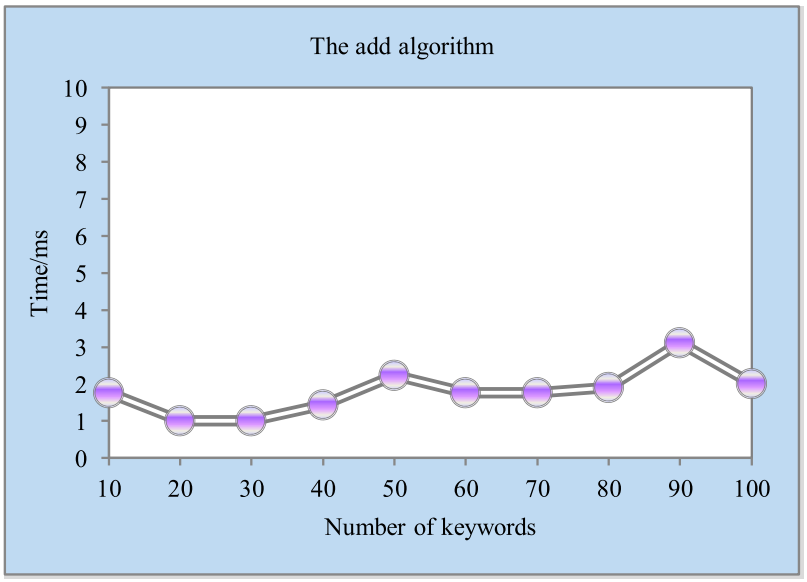

Figure 7. Add operation.

4. Research on computer big data processing

4.1. Research on the main overview of current cloud computing and big data

Big data refers to data that is difficult to process by conventional software and technology in a short time, so new processing methods and technologies are needed to process this part of the data. Among them, diversification refers to the diversification of information data formats and data types in the context of big data; and rapidity refers to the fact that in the context of big data, the transmission efficiency of information is higher, the processing speed is faster, and the information can be processed in a timely manner. Data is updated; while data review refers to the existence of some loopholes in the computer; the large amount of data refers to the large amount of big data computer information data processing, and it is also showing a trend of increasing day by day.

Both cloud computing and big data analysis need to be paid. Users can pay related fees according to their own needs, so as to obtain the resources actually needed. Data analysis is an important part of the big data processing process. The data is obtained and integrated and processed through related methods. In this process, the different values of the data are reflected. The combination of cloud computing network technology and big data can produce wonderful collisions: (1) In the cloud computing environment, different network users can obtain related resources according to their needs, which greatly extends the width of data information, and can also pass data Information access to network resources; (2) Improving the refinement of data analysis can deeply dig out the value of data, and it can also improve the application capabilities of software through cloud computing-based data analysis, thereby reducing the cost of data analysis. The advantages of combining data and cloud computing.

First of all, it is necessary to improve the level of data processing capabilities, which can help the system reflect the actual situation more objectively and comprehensively; on the other hand, it can also provide

a theoretical basis for decision makers. In addition, enhancing data processing capabilities can also dig deeper into the value of data and make it more cost-effective. Relevant departments and practitioners can dig deeper into the essence of data through research and data, and realize the sublimation of perceptual understanding of data.

4.2. Analysis of the basic processing flow of big data

Traditional data processing and input methods cannot meet the needs of massive data processing. If you want to process massive data, you must process and analyze the data in a short time, which requires more advanced information processing technology.

Big data processing is generally divided into four stages, which are as follows:(1) Data collection stage: With the increase in the number of Internet users, the amount of data inside the Internet has also begun to surge. In the face of such a large and complex data resource, how to collect data efficiently has become the key to big data processing.(2) Data processing and integration stage: The types involved in the data processing stage are more complex, and there are many redundant data that need to be deleted. Finally, the data with different formats is converted into a unified data format, so that the data processing can be more It's convenient and fast, and the most common processing method is a filter.(3) Data analysis stage: After the data is processed with a unified data structure, further analysis is required, and applications are classified according to the value of the data, and the data is processed centrally through various tools. At present, in the process of data analysis, there are already many products and software that specialize in big data analysis, which is of great help to the improvement of efficiency.(4) Data interpretation: Through this technology, the value of data and analysis results can be fully displayed to users, thereby improving the efficiency of users' application to the world and expanding the use value of data.

4.3. Analysis of the advantages and disadvantages of big data

The most prominent advantage of big data analysis is that it can visualize digitized information. In addition, big data data mining algorithms can enable relevant practitioners to mine the internal value of the data, thereby improving the cost-effectiveness of the data. In addition, big data analysis can also be applied to various fields for data prediction and data analysis. Data prediction technology is generally applied to fusion modeling technology, and new data fusion is carried out on the big data model.

Due to the increase in employees, the spread of data on social media has become more and more widespread. In this case, people's privacy is often leaked. Moreover, in a large amount of data, it is inevitable that some false or harmful information will be mixed, which not only affects the user's experience, but also triggers a series of uncontrollable events. Cloud computing network technology has four outstanding advantages, the specific analysis is as follows(1) Reduce computer costs;(2) Improve performance;(3) Almost unlimited data storage capacity;(4) Higher data storage security.

5. Conclusion

The rapid development of cloud technology has opened up a new path for data processing and storage. Due to its convenience and high efficiency, more and more enterprises and individual users have begun to store data in the cloud space. For some private data, most users choose to use encryption to ensure data security, but the encrypted data often fails to reflect the structural and semantic characteristics, which makes the encrypted documents unable to be retrieved by users in the cloud space. Keyword searchable encryption is the key technology to solve the above problems. It is a special encryption technology that can solve the retrieval problem after data encryption. The use of this technology allows users to find encrypted documents by searching for keywords, and ensures that intruders cannot obtain users' private information through keyword ciphertext searches or search credentials. This article has completed the following tasks by discussing the KSE program. (1) The existing scheme has been improved. The improved scheme can obtain the server through self-searching, so that the KSE scheme can resist the intrusion of bad external information, and does not need to use a secure channel. It can also be encrypted by combining attributes. Allows users to access the system through a variety of schemes. This solution has passed the four aspects of safety, correctness, functionality and performance tests, which fully proved the advantages and feasibility of the improvement; (2) Designed and constructed a cloud video sharing system, which combines the MU-DSSE solution it is applied to the storage scene of cloud video. With the development of information technology, the application of big data is becoming more and more extensive, which makes the processing technology of computer information gradually develop in the direction of informationization and large-scale. Therefore, more advanced and powerful technologies should be adopted to strengthen the computing power and storage capacity of the computer, so that the performance of the computer can keep up with the requirements of the development of the times. In the era of big data, the continuous increase of information data has brought certain difficulties to computer information processing. Therefore, it should be used for innovation, active improvement, and continuous enhancement of

computer performance, so that computer information processing capabilities can closely follow the development needs of the times.

Compliance with ethical standards

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Disclosure statement

The authors declare that they have no conflict of interests.

Data availability

Data will be made available on request.

ORCID

Lu Ming http://orcid.org/0009-0000-8945-9544

References

[1] Acampora G, et al. A survey on ambient intelligence in health care. In: Proc IEEE Inst Electr Electron Eng. 2013;101(12):2470–2494.
[2] Ballan L, et al. Event detection and recognition for semantic annotation of video. Multimed Tools Appl. 2010;51(1):279–302.
[3] Chan M, et al. A review of smart homes—present state and future challenges. Comput Methods Programs Biomed. 2008;91(1):55–81.
[4] Chatterjee N, Leuski A. A novel statistical approach for image and video retrieval and its adaption for active learning. Proceedings of the 23rd ACM International Conference on Multimedia. ACM, Brisbane. 2015;21(5):935–938.
[5] Chen L, Hoey J, et al. Sensor-based activity recognition. IEEE Trans Syst Man Cybern Part C (Appl Rev). 2012;15(2):1–19.
[6] Chen W, Ananthakrishnan S. ASR error detection in a conversational spoken language translation system. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2013;36(7):7418–7422.
[7] Chen L, Nugent CD, Wang H. A knowledge-driven approach to activity recognition in smart homes. IEEE Trans Knowl Data Eng. 2012;24(6):961–974.
[8] Choe TE, et al. Semantic video-to-video search using sub-graph grouping and matching. Proc IEEE Int Conf Comput Vis 1. 2013;57(6):787–794.
[9] Cook DJ, Das SK. How smart are our environments? An updated look at the state of the art. Pervasive Mobile Comput. 2007;3(2):53–73.
[10] De. Luca Ed, Bonacci S, Giraldi G. Aging populations: the health and quality of life of the elderly. La Clin Terapeut. 2011;162(1):e13.
[11] Filippova K, Hall K. Improved video categorization from text metadata and user comments. SIGIR'11 Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2011;23(8):835–842.
[12] Gaüzère B, et al. Semantic web technologies for object tracking and video analytics. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2015;94(7):574–585.
[13] Greco L, et al. Abnormal event recognition: a hybrid approach using semantic web. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2016;56(5):58–65.
[14] Hentschel C, Blümel I, Sack H. Automatic annotation of scientific video material based on visual concept detection. Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies, i-Know'. 2013;13(9):1–8.
[15] Hwang A, Hoey J. Smart home, the next generation: closing the gap between users and technology. AAAI Fall Symposium on Gerontechnology, Arlington. 2012;39(5):14–21.