Conference Proceeding Issue Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, www.ijtrd.com

Product Information Protect from Outsource Network

D. K. Gayathri
Master of Computer Science in Kamban College of Arts and Science For Women, Tamilnadu, India

Abstract: Cloud computing may be a promising info technique (IT) which will organize an outsized quantity of IT resources in associate economical and versatile manner. progressively various corporations arrange to move their native data management systems to the cloud and store and manage their product info on cloud servers. An related to challenge is a way to shield the safety of the commercially confidential knowledge, while maintaining the power to look the information. during this paper, a privacy-preserving knowledge search theme is planned, that can support each the identifier-based and feature-based product searches. Specifically, 2 novel index trees ar created and encrypted, which will be searched while not knowing the plaintext knowledge. Analysis and simulation results demonstrate the safety and potency of our theme.

Index Terms*: Product data retrieval, cloud computing, data security.*

I. INTRODUCTION

Driven by the revolution of knowledge technology in recent years and with the holdup within the economic process, there is associate pressing ought to rework China's entire industrial chain. To promote associate well-rounded industrial upgrading, China has proposed the strategy of ''Internet +'', and therefore the integration of China's ecommerce with its ancient economy has been considerably improved. Ecommerce has accelerated its expansion from consumption to numerous industries and infiltrated all aspects of social and economic activities, thereby driving the event of enterprise-level ecommerce, both in scope and full, and facilitating the transformation and upgrading of enterprises. The observation Report on the Data of China's Ecommerce Market [1] shows that in 2016, the volume of ecommerce transactions in China reached approximately three.5 trillion

bucks, a year-on-year rate of approximately twenty five.5%.

The chop-chop rising variety of cyber-transactions has spawned ecommerce massive knowledge. As progressively varied knowledge files square measure being hold on domestically in enterprises, the pressure on local knowledge storage systems greatly will increase. native hardware failures result in nice harm or loss of information, that greatly affects the daily operations of the enterprises. as luck would have it, cloud storage techniques came into being below such circumstances. Cloud computing will collect and organize a large number of various kinds of storage devices by means that of various functions, like cluster applications, network technology and distributed file systems. There have already been variety of typical cloud service merchandise reception and abroad, like Amazon net Services [2], Microsoft Azure [3], i Cloud [4], and App Engine [5].

As giant amounts of information square measure outsourced to cloud storage servers, the necessity for knowledge homeowners to encode the abovementioned second and third kinds of sensitive knowledge makes traditional plain text-based knowledge search solutions not suitable. additionally, restricted by the network information measure and local storage capability constraints, users notice it not possible to re-download all the info to a neighborhood disk and later decipher them to be used. supported the higher than problems, privacy-preserving data search schemes were born, designed to confirm that solely legitimate users supported identifiers or keywords, and have the ability to look the info. These schemes safeguard the users' personal knowledge however change the server to come back to the target ciphertext file in line with the question request. Thus, we can make sure the security of user knowledge and privacy whereas not unduly reducing the question potency.

In this paper, we tend to specialise in the second and third kinds of data and style a secure and economical knowledge search theme. For convenience, a sensible background is given as follows. We initial assume that every product contains a distinctive symbol in the whole company and an in depth description file. The file includes all of the careful data of the merchandise, such as the design flow, style customary, product options and market position. As we tend to all recognize, launching the merchandise to the market earlier than the competition will occupy the market quickly and benefit the corporate significantly. As a consequence, all of the information ought to be unbroken from the competitors and therefore the public, considering that the merchandise square measure time-sensitive.

With the expansion of the corporate, product data conjointly increases greatly. to boost the soundness and dependableness of a data storage system, associate intuitive theme is moving the local knowledge management system to the cloud. Cloud computing is wide treated as a promising data technique (IT) infrastructure owing to its powerful functionalities. It can collect and reorganize vast resources of storage, computing and applications, which suggests that the users will access the IT services in an exceedingly versatile, ubiquitous, economic and ondemand manner [10]. associate incidental challenge is a way to protect the confidentiality of the info whereas maintaining its searchability during this paper, we tend to style associate encrypted product information retrieval system. this method includes 2 index structures: a hash worth index tree, referred to as associate ID-AVL tree, and a height-balanced index tree, referred to as a product retrieval feature (PRF) tree. supported the 2 index trees, two knowledge search ways square measure supported, i.e., the info users can search the required product by the symbol or feature vector. the weather within the ID-AVL tree square measure the hash values of the product identifiers, instead of the plaintext knowledge, and the tree therefore is directly outsourced to the cloud. Meanwhile, the elements within the PRF tree square measure plaintext knowledge, and they are encrypted by the secure kNN formula before being outsourced. additionally, an in depth depth-first product search algorithm is intended for the PRF tree. Simulation results show the effectiveness and potency of the projected theme.

We summarize the first contributions of this paper as follows:

- A product data outsourcing and looking system model as well as the info owner, cloud server and data users is intended.
- 2 index structures supporting economical product retrieval square measure created. Moreover, corresponding search algorithms also are projected.
- we tend to integrate the secure kNN formula into our theme to guarantee the safety of the outsourced knowledge whereas maintaining the searchability.
- A series of simulations square measure conducted parenthetically the security and potency of the projected theme.

The rest of this paper is organized as follows. We first summarize the connected work of privacy-preserving knowledge search schemes in Section a pair of. Next, the info search system model and preliminary techniques square measure mentioned in Section three. Section four presents the encrypted product data retrieval scheme very well, and therefore the analysis of the projected theme is provided in Section five. Finally, the study's conclusions square measure presented in Section half dozen.

II. CONNECTED WORK

Cloud storage services have many blessings, such as ease of use and price saving, and that they square measure wide employed in many fields. However, many challenges square measure related to them. With the increasing quality of cloud storage, security problems became a crucial issue limiting its development. In recent years, knowledge discharge accidents have repeatedly occurred in such corporations as Microsoft, Google, Amazon, and China's Home hostel, Hanting, and Ctrip, and these incidents have exacerbated users' worries.

To counter the knowledge discharge, knowledge homeowners and enterprises generally source the encrypted business knowledge, rather than the plaintext knowledge, to cloud storage servers. In general, the outsourced knowledge is divided into 3 varieties. The first type is that the open-resource-type knowledge, that don't ought to be hidden from the cloud server, like the fundamental data of the enterprise and therefore the parameters of merchandise. The second type is that the non-public knowledge, which require to be encrypted however square measure only accessed and decrypted by the info contributor [6], [7].

This type includes such knowledge as internal counseling, intellectual properties and patents. The third kind is the non-public knowledge that require to be encrypted however also can be shared with specific users or teams [8], [9]. This type includes internal shared knowledge, hospital's division-wide case information {and data|and knowledge|and data} by some shared advanced users.

A single keyword Boolean search [7], [11]–[16] is that the simplest document retrieval technique for encrypted files. Song et al. [7] initial projected the searchable cryptography scheme during which every word in an exceedingly document is encrypted independently, and therefore the users ought to scan the complete document to look for an exact keyword. Consequently, this method has an especially high looking quality. Next, Goh [11] formally engineered the safety definitions for regular searchable cryptography, and a theme supported a Bloom filter was designed. the safety definitions square measure extended in [12] and [17]. because of the dearth of a rank mechanism for the came back results, the info users ought to take an extended time to screen the came back results, that is unacceptable generally. Thus, several single keyword-ranked search schemes have been projected [13]–[15], [18], [8]. tho' these schemes can come back a lot of correct search results, they can not satisfy users' needs in most cases, considering that one word cannot give ample data to explain the users' interests.

Multiple keyword Boolean search schemes enable the info users to input a group of keywords to look the required documents. Conjunctive keyword search schemes [19]–[21] come back the documents during which all the keywords given by the search question appear; divisional keyword search schemes return all the documents that contains a minimum of one keyword of interest. Predict keyword search schemes [22]–[24] have been projected to support each conjunctive and divisional search patterns. However, the came back results square measure still not sufficiently appropriate to the users as a result of the degrees of importance of the keywords aren't thought of in these schemes.

Cao et al. [25] initial projected a basic privacy-preservingmulti-keyword hierarchal search theme supported a secure kNN formula [26]. a group of strict privacy needs square measure established, and 2 schemes square measure later projected to boost the security and search expertise. However, an understandable downside of this theme is that the search potency is linear with the cardinality of the document assortment, and consequently, it can not be wont to method very large document databases. Xia et al. [27] designed a keyword balanced binary tree to prepare the document vectors and projected a ''Greedy Depth-First Search'' formula to improve the search potency. Moreover, the index tree will be updated dynamically with an appropriate communication burden. Chen et al. [28] took the relationships of documents into thought, and a hierarchical-clustering-based index structure was designed to boost the search potency. In addition, a verification theme was conjointly integrated into their theme to ensure the correctness of the results. However, these 2 index trees in [27] and [28] is additional improved in terms of potency and accuracy as mentioned in Section one. Fu et al. [29] given a personalised multikeyword hierarchal search theme during which associate interest model of the users is integrated into the document retrieval system to support a personalised search and improve the users' search experience. Specifically, the interest model of an information user is built supported his search history with the assistance of WordNet [30] so as to depict his behaviors in fine grit level. However, this theme doesn't support dynamic update operations as a result of the document vectors square measure created based mostly on all the documents. additionally, tho' associate MDB-tree is employed to boost the search potency, the effectiveness of the tree is troublesome to predict. many alternative connected studies in the field of cloud computing is found in [33]–[37].

III. SYSTEM MODEL AND THEREFORE THE INDEX TREES

A. Product Retrieval System

As shown in Fig. 1, the complete product retrieval system model is composed primarily of 3 entities: the info manager, the cloud server and therefore the knowledge user. the first responsibilities of those 3 entities square measure given within the following.

The data manager is to blame for managing the merchandise and grouping the merchandise data. additionally, the data manager has to encode the merchandise data file by a

symmetric cryptography technique before outsourcing the info to the cloud server. to boost the safety of the files, each file is encrypted by one secret key, and therefore the keys of different files square measure freelance. what is more, to improve the search potency, associate index structure is built for the outsourced knowledge. At first, associate symbol index structure is constructed supported the hash operate and height-balanced binary search tree. Then, a feature vector tree is constructed for all the feature vectors of the merchandise, and it's encrypted by the secure kNN formula.

When an information user needs to look a group of chosen merchandise, she has to generate a trapdoor to explain her interest. Two kinds of the trapdoor is provided, i.e., a set of

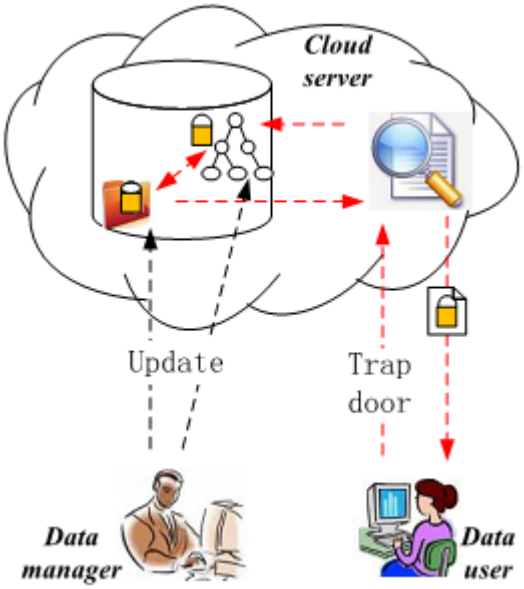

FIGURE 1. Encrypted product data retrieval system model.

hash values of the required product data files or a group of feature vectors. For the primary sort of trapdoor, a set of encrypted files with identical hash identifiers square measure came back, and for the second kind trapdoor, the foremost relevant encrypted files square measure came back. the info user will acquire the plaintext files by decrypting the came back files with the assistance of the symmetric secret keys. These secret keys square measure provided by the data manager.

The cloud server stores all the info uploaded by the info manager. once an information user has to search the info within the cloud, she initial generates a trapdoor, that is distributed to the cloud server. a quest engineer is utilized by the cloud server to act as a bridge between {the knowledge|the info|the information} users and therefore the encrypted data. Though the cloud server cannot get the plaintexts of the info, it ought to be capable of causing the correct search results of the trapdoor to the info users. Of course, the came back knowledge are ciphertext, and therefore the knowledge user has to decipher them by the regular secret keys that square measure provided by the info manager.

B. ID-AVL Tree

To construct the ID-AVL tree, we tend to initial encode all the merchandise identifiers supported a hash operate, hash(). Next, each node in the ID-AVL tree contains a hash worth of the merchandise ID, and all of the hash values square measure organized supported associate AVL tree [31] as shown in Fig. 2. 2 necessary properties of AVL, which might facilitate U.S. to take care of the hash values, are presented as follows. First, the ID-AVL tree is updated flexibly by inserting a node, deleting a node and modifying a node. Correspondingly, we are able to update the ID-AVL tree from time to time by dynamic the merchandise data. Second, the values of the left kid nodes of a parent node square measure perpetually smaller than that of the parent node; the values of the correct child nodes of a parent node square measure perpetually larger than that of the parent node. In theory, the time quality of inserting, deleting and looking a node square measure all log(N), wherever N is that the number of nodes within the tree. during this paper, we tend to construct the ID-AVL tree supported the formula in [31].

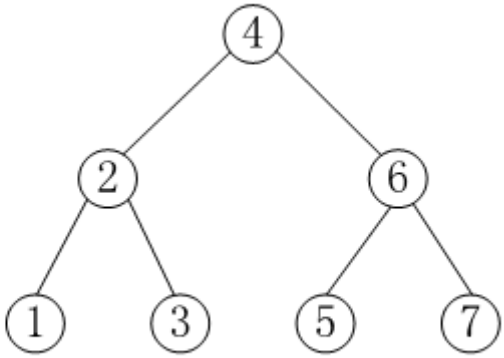

FIGURE 2. Product hash value index tree.

FIGURE 2. Product hash worth index tree.

C. Product Retrieval Tree

The feature wordbook of the merchandise is denoted as D = $\{f_1, f_2, \cdots, f_m\}$, and therefore the feature set Si of any product Pi should be a set of D, i.e., i.e., $S_i \in 2^{\{f_1, f_2, \cdots, f_m\}}$. Then, the feature vector Vi of product Pi is constructed as follows.

• The initial vector of Pi is a one × m vector, and every one the elements within the vector square measure 0;

• we tend to orderly scan all the weather within the initial vector and assign a price to the component of feature fi if the feature of Pi is quantity.

• supported the various degrees of importance of the features, a weight is utilized to multiply the weather in the vector to replicate this.

To search the merchandise data, a trapdoor has to be constructed by the info user in an exceedingly similar means, and therefore the similarities between the trapdoor VQ and therefore the product feature Vi is calculated as sim

Vi, VQ= Vi·VQ. Moreover, the similarity

between 2 the vectors Vi and Vj is outlined as sim

$(V_i, V_Q) = V_i \cdot V_Q$. Next, the merchandise feature vectors square measure organized as $(V_i, V_j) = V_i \cdot V_j$ hierarchical clusters in line with their similarities. Each node in the tree represents a cluster composed of a group of product feature vectors or sub-clusters. The PRF vector of a node could be a quintuple account a couple of cluster.

Given K m-dimensional product feature vectors in an exceedingly cluster:

Vj where j = one, 2, · · · ,K, the PRF vector of the cluster is denoted as a quintuple: PRF = (K, LS, SS, Vmin, Vmax), where K is that the variety of product feature vectors within the cluster, LS is that the linear total of the K product feature vectors,

$$\text{i.e., } LS = \sum_{j=1}^{K} V_j,$$

SS is that the sq. total of the K product feature vectors $\text{i.e., } SS = \sum_{j=1}^{K} V_j^2$ (SS could be a numerical worth rather than a vector), Vmin denotes a vector consisting of m values that square measure calculated as follows:

$$V_{min}[i] = min(V_1[i], V_2[i], \cdots, V_K[i]), \quad (2)$$

where Vj[i] is that the i-th dimensional worth of Vj , and equally, Vmax is calculated as follows:

$$V_{max}[i] = max(V_1[i], V_2[i], \cdots, V_K[i]). \quad (3)$$

Based on a PRF vector, the centre of mass of a cluster C is easily calculated as

$$c = LS/K, \quad (3)$$

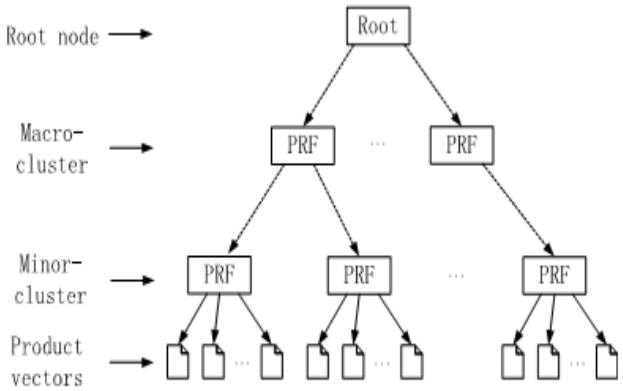

FIGURE 3. Product retrieval feature tree.

FIGURE 3. Product retrieval feature tree.

and the connection score between cluster C and a product vector Vj is outlined as

$$\text{RScore}(C, V_j) = c \cdot V_j. \quad (4)$$

Similarly, the connection score between cluster C and a question vector VQ is outlined as

$$\text{RScore}(C, V_Q) = c \cdot V_Q. \quad (5)$$

Further, the radius of cluster C is outlined as follows:

$$R = \sqrt{\sum_{j=1}^{K} (V_j - c)^2 / K}, \quad (6)$$

and it is calculated by the PRF vector as follows:

$$R = \sqrt{(SS - LS^2/K)/K}. \quad (7)$$

Theorem one (PRF Additivity Theorem): If we tend to merge two disjoint clusters with PRF vectors: $PRF_1 = (K_1, LS_1, SS_1, V_{min1}, V_{max1})$ and $PRF_2 = (K_2, LS_2, SS_2, V_{min2}, V_{max2})$, then the PRF vector of the combined cluster is

$$PRF = PRF_1 + PRF_2 = (K_1 + K_2, LS_1 + LS_2, SS_1 + SS_2, V_{min}, V_{max}), \quad (8)$$

where $V_{min}[i] = \min(V_{min1}[i], V_{min2}[i])$ and $V_{max}[i] = max(V_{max1}[i], V_{max2}[i])$.

Proof: The proof consists of simple pure mathematics. In the similar means, we are able to acquire the PRF Subtraction Theorem, which might be wont to divide 2 clusters, though Vmin and Vmax ought to be recalculated. The structure of a PRF tree is given in Fig. 3. It is discovered that every leaf node consists of a group of comparable product vectors and its PRF vector is directly extracted from the merchandise vectors. The similar leaf nodes agglomerate with one another to compose the non-leaf nodes till all the merchandise vectors belong to a huge cluster at the a root node. supported Theorem one, the PRF vectors of the non-leaf nodes and therefore the root node square measure calculated based on the PRF vectors of all their kid nodes.

IV. ENCRYPTED PRODUCT DATA RETRIEVAL THEME

A. Construction of Product Retrieval Tree

A PRF tree has 3 main parameters: branching factors B1, and B2 and threshold T , that square measure predetermined by the info owner. every non-leaf node NLi contains at the most B1 kid nodes, and it's outlined as follows:

$$NL_i = (PRF, PRF_1, child_1, \cdots, PRF_{B_1}, child_{B_1}) \quad (9)$$

where PRF is that the PRF vector of the complete cluster, PRFi is the PRF vector of the i-th sub-cluster and childi is a pointer to the kid node representing the sub-cluster. A non-leaf node represents a cluster created of all the sub-clusters represented by its kid nodes. A leaf node Li contains at the most B2 product vectors, and it's outlined as follows:

$$L_i = (PRF, child_1, \cdots, child_{B_2}), \quad (10)$$

where PRF is that the PRF vector of the cluster, childi is a pointer to the i-th product vector within the cluster. what is more, the cluster of a leaf node should satisfy a threshold requirement: the radius of the cluster (11) should be but T . The default values within the nodes square measure set to null.

The PRF tree is built in associate progressive manner, and the process of inserting a product vector Vj into the PRF tree is given as follows:

• distinctive the acceptable leaf node: ranging from the root, Vj recursively descends the PRF tree by selecting the nighest kid node in line with the connection scores between Vj and therefore the sub-clusters as outlined in (11) till it reaches a leaf node.

• Modifying the leaf node: once Vj reaches a leaf node Li , it tests whether or not Li will ‘‘absorb’’ Vj while not violating the constraints of B2 and T. If so, Vj is inserted into Li and the PRF vector of Li is updated. If not, we must split Li to two leaf nodes. Node ripping is performed by selecting the farthest try of product vectors as seeds and redistributing the remaining

product vectors based mostly on the nighest criteria. The PRF vectors of the 2 new leaf nodes ought to be recalculated.

• Modifying the trail from the basis node to the leaf node: After inserting Vj into a leaf node, we want to update the PRF vector for all the nodes on the trail to the leaf node. within the absence of a split, this merely involves updating PRF vectors supported Theorem one. A leaf node split needs U.S. to insert a replacement leaf node into the parent node. If the parent node has house for the new leaf node, we solely ought to insert the new leaf node into it then update the PRF vector for the parent node. In general, however, we tend to might need to split the parent node furthermore, and so on, up to the basis. If the basis is split, the tree height increases by one.

B. Retrieval Method of The Interested Merchandise

In this paper, the info users will retrieve the interested product in 2 ways in which, i.e., retrieving the merchandises by their identifiers or the product feature vector. once an information user needs to search a product supported its symbol, she initial has to encrypt the symbol based mostly the on the hash operate, hash(). Next, the hash worth of the symbol is distributed to the cloud server. The cloud server is to blame for sorting out the hash worth within the ID-AVL tree, and once the hash worth is found, the corresponding encrypted production data is sent to the info user. Finally, the info user will decipher the product data supported the key keys, and therefore the knowledge retrieval method is completed.

Moreover, in bound cases, the info user might want to look the product supported the options. Initially, the info user desires to construct the feature vector of the merchandise as mentioned in Section three.3.

Then, we want to style a depth-first search algorithm for the PRF tree, which formula is given in Algorithm one.

Algorithm 1 DepthFirstSearch(a PRF Tree With Root r, a Query Vector V_Q)

```
1:  u ← r;
2:  while u is not a leaf node
3:    Calculate all the relevance scores between the child
        nodes of u with V_Q based on (5);
4:  u ← the most relevant child node;
5:  end while
6:  Select the most relevant k document vectors in u by
    RScore(V_i, V_Q) and construct RList;
7:  Stack.push(r);
8:  while Stack is not empty
9:   u ← Stack.pop();
10:  if the node u is not a leaf node
11:   if RScore (V_{u,max}, V_Q) > kthScore
12:    Sort the child nodes of u in ascending order based
         on the relevant scores with V_Q;
13:    Push the children of u into Stack in order, i.e., the
         most relevant child is latest inserted into Stack;
14:   else
15:    break;
16:   end if
17:  else
18:   Calculate the relevance scores between the document
        vectors in the leaf node with V_Q and update RList;
19:  end if
20: end while
21: return RList;
```

In formula one, the kthScore represents the littlest connection score within the current result list RList, that stores the most k relevant accessed document vectors with VQ and therefore the corresponding connection scores. additionally, we tend to use the variable Stack to store the nodes which require to be searched in the future. additionally, Stack.push(u) inserts node u into Stack and Stack.pop() returns the most recent inserted node. within the initial phase, we want to initial find the foremost relevant leaf node with the question vector within the tree to initialize RList as given in line a pair of to line half dozen. Then, the result list is endlessly updated by looking the required ways within the tree till the ultimate search result's obtained as given in line eight to line nineteen. Compared with the search method of the keyword balanced binary tree projected in [27], the search method given in

Algorithm one is far a lot of economical considering that several search ways square measure cropped within the looking method.

C. cryptography OF the merchandise RETRIEVAL TREE

For each product Pi , 2 kinds of data square measure initial extracted, as well as its symbol i and therefore the product vector Vi We encode the symbol i through a hash operate, hash(). The construction method of the ID − AVL tree is given as follows. The created ID − AVL tree is directly outsourced to the cloud server as a result of it stores solely a group of hash values, instead of the plaintext symbol.

Based on the merchandise vectors, the method of building thePRF tree has been given in Section four.2. In distinction to the ID − AVL, the PRF tree has to be encrypted before being outsourced. within the PRF tree, we tend to treat LS, Vmin and Vmax to the same as product vectors and encode them in identical way. Note that parameter K in an exceedingly PRF vector doesn't would like to be encrypted, and SS, which can not be employed in the search process, doesn't ought to be sent to the cloud server. Before

encrypting a product vector Vj in the PRF tree, we tend to initial extend it to (m+m0) dimensions. additionally, we tend to spilt every dimension of Vj[i] into Vj[i] 0 and Vj[i]00. Specifically, if S2i = 0, Vj[i]0 and Vj[i]00 are set adequate Vj[i]; otherwise, Vj[i]0 and Vj[i]00 are set as 2 random numbers whose total is equal to Vj[i]. Next, we tend to indiscriminately choose 2 invertible matrices M1, M2 and encode Vj as $E_j = \{M_1^T V_j', M_2^T V_j''\}$.

Once a quest request SR is received by the proxy server, it initial extracts its parameters as well as ID0 and vSR. Parameter ID0 is encrypted by hash() and that we get hID0 . We extend vSR to (m + m0) dimensions. Specifically, if S1i = 0, the i-th dimension of VQ corresponds to a feature wr, which is extracted from W so as, and VQ[i] is ready to wwr ; otherwise, this dimension is a man-made dimension and VQ[i] is set to a random variety. Note that the worth of the last artificial dimension isn't a random variety, and it ought to be calculated rigorously to ensure that the scalar product of the by artificial means additional dimensions within the product vectors and in VQ is 0. Further, we tend to spilt VQ[i] into VQ[i]0 and VQ[i]00 Specifically, if S2i = 1, VQ[i]0 and VQ[i]00 are set adequate VQ[i]; otherwise, VQ[i]0 and VQ[i]00 are set as 2 random numbers whose total is adequate VQ[i]. Finally, we encrypt VQ as combining weight = . during this case, the connection score of Vj and VQ outlined in Section three.2 is calculated as follows:

$$\text{RScore}\left(V_j, V_Q\right) = V_j \cdot V_Q = E_j \cdot E_Q. \qquad (11)$$

The trapdoor TD consists of the hash values of the filename and authors and combining weight.

V. PERFORMANCE ANALYSIS

A. Security Analysis

In our theme, the outsourced knowledge includes the merchandise information file, ID-AVL tree and PRF tree. the merchandise information files square measure encrypted symmetrically supported the independent secret keys, and therefore the cloud server doesn't have

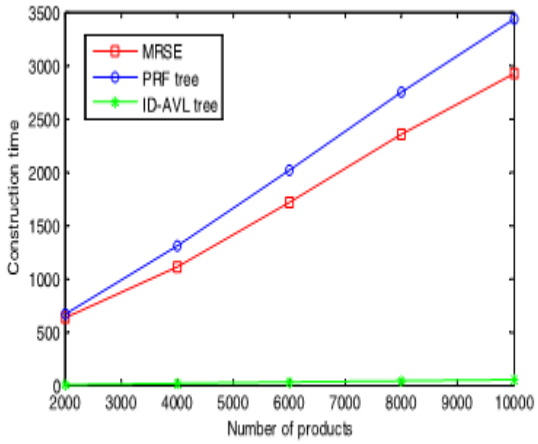

FIGURE 4. Construction time of the index structures.

Figure 4: Construction Time Of The Index Structures.

the secret keys. during this case, the plaintext files can not be decrypted by the cloud server. within the ID-AVL tree, the stored values square measure the hash values of the merchandise identifiers, and they contain no valuable data concerning merchandise. The PRF tree is encrypted by the secure kNN formula before being outsourced to the cloud server. tho' the cloud server is aware of the encrypted feature vectors within the tree, the cloud server will not recognize the matrices M1, M2; hence, the plaintext vectors in the tree can not be recovered.

B. PRODUCT data SEARCH potency

In this section, we tend to measure the search potency of our scheme. First, we tend to measure the development time of the index structures of the merchandise data. Specifically, we tend to compare our theme with the MRSE theme [25]. To decrease the bias of the info manager WHO is to blame for generating the vectors and therefore the hash values, during this paper we tend to use the Enron Email knowledge Set [32] to check our theme. Specifically, the data set is utilized to act because the product data files. Moreover, the vectors of the merchandise square measure assumed to be extracted from the info set supported the TF-IDF model, and then the vectors square measure organized by the PRF tree. As shown in Fig. 4, with the increasing variety of merchandise, the construction time of PRF tree and therefore the index structures in MRSE monotonously increase. this is often affordable considering that every product data file has to be scanned for a time to induce the feature vectors. the development time of the PRF tree is slightly longer than that of the MRSE theme, because the vectors ought to be additional inserted to the trees in the PRF tree. Apparently, the ID-AVL tree is significantly simpler, and therefore the construction time is unheeded compared with the opposite 2 trees.

To search the required product data, the info user needs to initial generate the trapdoor, that is distributed to the cloud server. the days of constructing the trapdoors with the increasing of the dimensions of the feature wordbook square measure given in Fig. 5. The search requests supported the identifiers square measure independent of the feature wordbook, and hence, the time of constructing the trapdoors for the ID-AVL tree remains stable. However, the development time of the trapdoors for the MRSE and PRF trees monotonously increase with the

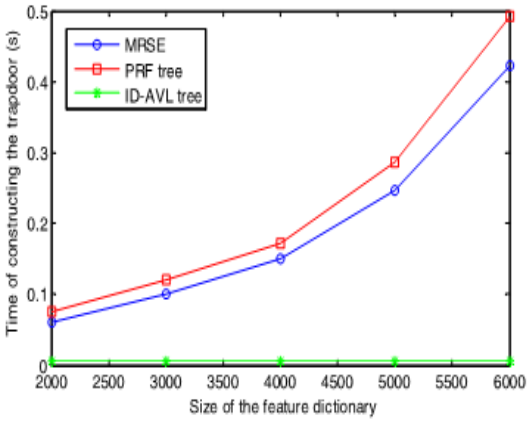

FIGURE 5. Time of constructing the trapdoors.

Figure 5: Time of constructing the trapdoors.

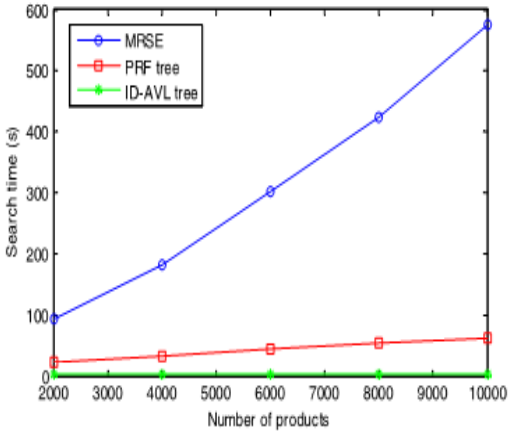

FIGURE 6. Search time with different number of products.

Figure 6: Search time with totally different variety of merchandise.

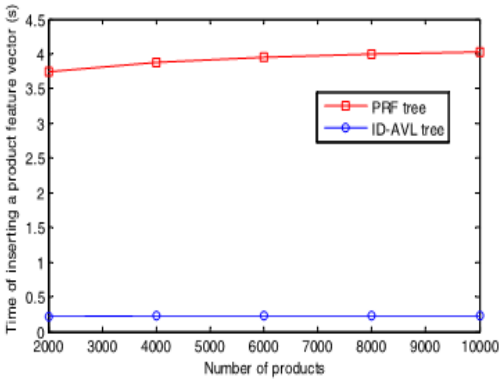

FIGURE 7. Time consumption of inserting a node into the trees.

Figure 7: Time consumption of inserting a node into the trees.

increasing of the feature dictionary's size. this is often affordable considering that the dimensions of the merchandise feature vector is equal to the dimensions of the feature wordbook. additionally, the time prices for the MRSE and PRF trees square measure almost like one another as a result of the processes of generating the trapdoors square measure similar.

The search time of a trapdoor within the cloud server is given in Fig. 6. It is discovered that the MRSE theme consumes the foremost time to execute a quest operation. Moreover, the search time will increase monotonously with the increasing of the amount of merchandise. This increase is explained by the very fact that in MRSE, the feature vectors square measure stored so as, and that they don't use any index structure.

In this case, the cloud server has to scan all the merchandise feature vectors to induce the search result. The PRF tree organizes the vectors by a height-balanced tree, and most ways within the tree square measure cropped within the search method. As a consequence, the search potency is greatly improved. Finally, we can observe that the ID-AVL tree is that the best index structure, which might be explained by the very fact that the ID-AVL tree is significantly easier, and therefore the search method is additionallyvery easy.

With the increasing growth of corporations, a lot of and a lot of product data has to be outsourced to the cloud server. Consequently, we want to update the index trees from time to time, {and the|and therefore the|and conjointly the} update potency also affects the performance of our theme considerably. As shown in Fig. 7, the update time of each the PRF tree and therefore the ID-AVL tree increases slightly with the increasing variety of merchandise, which is cheap, considering that we want to look the trees to spot the correct location of the inserted node. In addition, change the PRF tree consumes way more energy than that of change the ID-AVL tree. this may be explained by the very fact that the ID-AVL tree is far easier than the PRF tree, and in theory solely log(N) nodes would like to be searched. tho' quite several ways within the PRF tree are cropped within the search method, the amount of the search paths is significantly larger than log(N) and longer is therefore consumed within the PRF tree.

CONCLUSIONS

In this paper, we tend to designed a secure and economical product data retrieval theme supported cloud computing. Specifically, 2 index structures, as well as a hash value AVL tree and a product vector retrieval tree, square measure created, and that they support associate identifier-based product search and feature-vector-based product search, severally. Correspondingly, 2 search algorithms square measure designed to look the two trees. to guard the merchandise data privacy, all the outsourced knowledge square measure encrypted. the merchandise data is symmetrically encrypted supported a group of freelance secret keys, and therefore the product vectors square measure encrypted supported the secure kNN formula. Security analysis and simulation results illustrate the safety and potency of the projected scheme. As the future work, we tend to decide to seamlessly integrate more index structures into our theme to support a lot of search patterns. Another troublesome and promising challenge is additional improving the search potency.

References

[1] (May 24, 2017). 2016 Monitoring Report on the Data of China's E-Commerce Market [EB/OL]. [Online]. Available: http://www.100ec. cn/zt/16jcbg/

[2] Amazon. Amazon S3. Accessed: Sep. 5, 2017. [Online]. Available: http://aws.amazon.com/s3/

[3] Windows Azure. Accessed: Sep. 5, 2017. [Online]. Available: http://www.microsoft.com/windowsazure/

[4] Apple i Cloud. Accessed: Sep. 5, 2017. [Online]. Available: http://www.icloud.com/ [5] Google App Engine. Accessed: Sep. 5, 2017. [Online]. Available: http://appengine.google.com

[5] P. Golle, J. Staddon, and B. Waters, ''Secure conjunctive keyword search over encrypted data,'' in Proc. Int. Conf. Appl. Cryptogr. Netw. Secur., 2004, pp. 31–45.

[6] D. X. Song and D. A. Wanger Perrig, ''Practical techniques for searches on encrypted data,'' in Proc. IEEE Symp. Security Privacy, May 2000, pp. 44–55.

[7] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, ''Public key encryption with keyword search,'' in Proc. Int. Conf. Theory Appl. Cryptogr. Techn., 2004, pp. 506–522.

[8] H. S. Rhee, J. H. Park, W. Susilo, and D. H. Lee, ''Trapdoor security in a searchablepublic-keyencryptionschemewithadesignatedtester,''J.Syst. Softw., vol. 83, no. 5, pp. 763–771, 2010.

[9] K.Ren,C.Wang,andQ.Wang,''Securitychallengesforthe publiccloud,'' IEEE Internet Comput., vol. 16, no. 1, pp. 69–73, Jan./Feb. 2012.

[10] E.-J. Goh, ''Secure indexes,'' IACR Cryptol. ePrint Arch., Newark, NJ, USA, Tech. Rep. 1, 2003, p. 216.

[11] R.Curtmola,J.Garay,S.Kamara,andR.Ostrovsky,''Searc hablesymmet- ric encryption: Improved definitions and efficient constructions,'' J. Com- put. Secur., vol. 19, no. 5, pp. 895–934, 2011.

[12] A. Swaminathan et al., ''Confidentiality-preserving rank-ordered search,'' in Proc. ACM Workshop Storage Secur. Survivability, 2007, pp. 7–12.

[13] C. Wang, N. Cao, K. Ren, and W. Lou, ''Enabling secure and efficient ranked keyword search over outsourced cloud data,'' IEEE Trans. Parallel Distrib. Syst., vol. 23, no. 8, pp. 1467–1479, Aug. 2012.

[14] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, ''Zerber+R: Top-k retrieval from a confidential index,'' in Proc. 12th Int. Conf. Extending Database Technol., Adv. Database Technol., 2009, pp. 439–449.

[15] S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, ''Outsourced symmetric private information retrieval,'' in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2013, pp. 875–888.

[16] Y. C. Changand M.Mitzenmacher, ''Privacy preserving keyword searches on remote encrypted data,'' in Proc. Int. Conf. Appl. Cryptogr. Network Secur., 2005, pp. 442–455.

[17] C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, ''Secure ranked keyword search over encrypted cloud data,'' in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2010, pp. 253–262.

[18] L. Ballard, S. Kamara, and F. Monrose, ''Achieving efficient conjunctive keyword searches over encrypted data,'' in Proc. Int. Conf. Inf. Commun. Secur., 2005, pp. 414–426.

[19] Y. Ho Hwang and P. J. Lee, ''Public key encryption with conjunctive keyword search and its extension to a multi-user system,'' in Proc. Int. Conf. Pairing-Based Cryptogr., 2007, pp. 2–22.

[20] B. Zhang and F. Zhang, ''An efficient public key encryption with conjunctive-subset keywords search,''

J. Netw. Comput. Appl., vol. 34, no. 1, pp. 262–267, 2011.

[21] A. Lewko, T. Okamoto, A. Sahai, K. Takashima, and B. Waters, ‘‘Fully secure functional encryption: Attribute-based encryption and (hierarchical) innerproductencryption,’’inProc.Annu.Int.Conf.Theory Appl.Cryptogr. Techn., 2010, pp. 62–91.

[22] E.Shen,E.Shi,andB.Waters,‘‘Predicateprivacyinencryptionsystems,’’ in Proc. Theory Cryptogr. Conf., 2009, pp. 457–473.

[23] J.Katz,A.Sahai,andB.Waters,‘‘Predicateencryptionsupportingdisjunc-tions,polynomialequations,andinnerproducts,’’inProc. Annu.Int.Conf. Theory Appl. Cryptogr. Techn., 2008, pp. 146–162.

[24] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, ‘‘Privacy-preserving multi- keyword ranked search over encrypted cloud data,’’ IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2014.

[25] W. K. Wong, D. W. L. Cheung, B. Kao, and N. Mamoulis, ‘‘Secure knn computation on encrypted databases,’’ in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2009, pp. 139–152.

[26] Z. Xiam, X. Wang, X. Sun, and Q. Wang, ‘‘A secure and dynamic multi- keyword ranked search scheme over encrypted cloud data,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 2, pp. 340–352, Feb. 2016.

[27] C. Chen et al., ‘‘An efficient privacy-preserving ranked keyword search method,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 4, pp. 951–963, Apr. 2016.

[28] Z.Fu,K.Ren,J.Shu,X.Sun,andF.Huang,‘‘Enablingpersonalizedsearch over encrypted outsourced data with efficiency improvement,’’ IEEE Trans. Parallel Distrib. Syst., vol. 27, no. 9, pp. 2546–2559, Sep. 2016.

[29] G. A. Miller, ‘‘WordNet: A lexical database for English,’’ Commun. ACM vol. 38, no. 11, pp. 39–41, 1995.

[30] G. Adelson-Velsky and G. E. Landis, ‘‘An algorithm for the organization of information,’’ (in Russian), in Proc. USSR Acad. Sci., vol. 146. 1962, pp. 263–266.

[31] W. W. Cohen. (2015). Enron Email Data Set. [Online]. Available: https://www.cs.cmu.edu/~./enron/

[32] C. Zhu, J. J. P. C. Rodrigues, V. C. M. Leung, L. Shu, and L. T. Yang, ‘‘Trust-based communication for the industrial Internet of Things,’’ IEEE Commun. Mag., vol. 56, no. 2, pp. 16–22, Feb. 2018.

[33] C.Zhu,L.Shu,V.C.M.Leung,S.Guo,Y.Zhang,andL.T.Yang,‘‘Secure multimedia big data in trust-assisted sensor-cloud for smart city,’’ IEEE Commun. Mag., vol. 55, no. 12, pp. 24–30, Dec. 2017.

[34] C. Zhu, H. Zhou, V. C. M. Leung, K. Wang, Y. Zhang, and L. T. Yang, ‘‘Toward big data in green city,’’ IEEE Commun. Mag., vol. 55, no. 11, pp. 14–18, Nov. 2017.

[35] C. Zhu, V. C. M. Leung, K. Wang, L. T. Yang, and Y. Zhang, ‘‘Multi-method data delivery for green sensor-cloud,’’ IEEE Commun. Mag., vol. 55, no. 5, pp. 176–182, May 2017.