



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 2, February 2021

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.488

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

An Efficient and Privacy-Preserving Multi Keyword Ranked Search over Encrypted Data

Monika D. Rokade, Nutan S. Shelke

Assistant Professor, Dept. of Computer, SPCOE, (Dumbarwadi), Otur, Pune , Savitribai Phule Pune University, Pune, Maharashtra, India

PG Student, Dept. of Computer, SPCOE, (Dumbarwadi), Otur, Pune , Savitribai Phule Pune University, Pune, Maharashtra, India

ABSTRACT: Data Mining has wide applications in numerous zones, for example, keeping money, prescription, investigative exploration and among government offices. Order is one of the ordinarily utilized assignments as a part of information mining applications. For as far back as decade, due to the ascent of different protection issues, numerous hypothetical and commonsense answers for the order issue have been proposed under diverse security models. Notwithstanding, with the late fame of distributed computing, clients now have the chance to outsource their information, in encoded structure, and also the information mining assignments to the cloud. Since the information on the cloud is in encoded structure, existing security protecting characterization methods are not appropriate. In this paper, system concentrates on fathoming the characterization issue over encoded information. Specifically, system proposes a safe k-NN classifier over scrambled information in the cloud. The index is created with the help of Vector base cosine similarity (VCS) multiple strings matching algorithm which matches the pre-defined set of keywords with information in the data files to index them and store relevant data. The proposed convention ensures the classification of information, security of client's data inquiry, and shrouds the information access designs. To the best of our learning, our work is the first to add to a safe k-NN classifier over scrambled information under the semi-legitimate model. Additionally, system exactly dissects the effectiveness of our proposed convention utilizing a genuine dataset under diverse parameter settings.

KEYWORDS: Security, k-NN classifier, outsourced databases, encryption data.

I. INTRODUCTION

Lately, the cloud computing model [1] is changing the landscape of the organizations way of working their information especially in the way they save access and process data. As a growing processing model, cloud processing draws many organizations to think about seriously concerning cloud potential with regards to its cost-efficiency, versatility, and offload of management expense. Most often, organizations assign their computational functions in improvement to their information to the cloud. Regardless of remarkable benefits that the cloud offers, security and comfort issues in the reasoning are avoiding companies to utilize those benefits. When information is extremely delicate, the information need to be encoded before freelancing to the cloud. Nevertheless, when information are secured, regardless of the actual security plan, executing any information mining tasks turns into very complicated without ever decrypting the information.

There are other privacy worries, confirmed by the following example. Assume an insurance provider contracted its secured clients database and relevant data mining task to a cloud. When a representative from the company needs to figure out the threat stage of a potential new client, the representative can use a classification method to figure out the threat stage of the client. Initial, the representative requires generating a details history q for the client containing certain private details of the client, e.g., credit rating, age, marriage status, etc. Then this history can be sent to the cloud, and the cloud will estimate the class label for q . However, since q contains vulnerable details, to secure the customer's privacy, q should be encoded before delivering it to the cloud.

The above example reveals that data mining over encoded information on a cloud also requires securing a user's history when the history is a part of a data mining procedure. Furthermore, cloud can also obtain helpful and delicate information about the real information products by monitoring the information accessibility styles even if the information are encoded.

II. RELATED WORK

In [1] Author discussed hierarchical approach clusters the documents based on the minimum relevance threshold, and then partitions the resulting clusters into sub-clusters until the constraint on the maximum size of cluster is reached. In the search phase, this approach can reach a linear computational complexity against an exponential size increase of document collection. In order to verify the authenticity of search results, a structure called minimum hash sub-tree is designed in this approach. System also investigated cipher text search in the scenario of cloud storage. System explore the problem of maintaining the semantic relationship between different plain documents over the related encrypted documents and give the design method to enhance the performance of the semantic search. In [2] Author System proposed a protocol of finding frequent item in accountable computing (AC) framework which enables two parties to conduct collaborative computation on their transactional databases to find out the common frequent items without disclosing their private data to the other party. Their scheme was proposed in a secure two-party computation model against malicious adversaries. System also analyses the implementation details of AC-framework and identifies some security weaknesses in their scheme. Furthermore, system clarifies the security requirements for the AC-framework and presents an augmented solution to enhance security. System also analyses the search efficiency and security under two popular threat models. In [3] Author proposed an efficient and flexible protocol, called EFPA, for privacy-preserving association rule mining in cloud. With the protocol, plenty of participants can provide their data and mine the association rules in cloud together without privacy leakage. Detailed security analysis shows that the proposed EFPA protocol can achieve privacy-preserving mining of association rules in cloud. It also present an efficient and flexible privacy preserving association rule mining protocol, called EFPA. Unlike most existing works, EFPA can support distributed data providers to collaboratively achieve association rule mining without exposing any privacy of data providers or mining results, i.e, the providers' data and mining results cannot be revealed by cloud. In [4] Author proposed k-NN protocol protects the confidentiality of the data, user's input query, and data access patterns. To the best of this knowledge, this work is the first to develop a secure k-NN classifier over encrypted data under the standard semi-honest model. In this system, author focus on solving the classification problem over encrypted data. In particular, propose a secure k-NN classifier over encrypted data in the cloud. The proposed protocol protects the confidentiality of data, privacy of user's input query, and hides the data access patterns. In [5] Author focus on cloud-aided frequent itemset mining solution, which is used to build an association rule mining solution. Here outsourced databases that allow multiple data owners to efficiently share their data securely without compromising on data privacy and leak less information about the raw data than most existing solutions. In comparison to the only known solution achieving a similar privacy level as this proposed solutions, the performance of this proposed solutions is three to five orders of magnitude higher. Based on this experiment findings using different parameters and data sets, system demonstrate that the run time in each of this solutions is only one order higher than that in the best non-privacy-preserving data mining algorithms. Since both data and computing work are outsourced to the cloud servers, the resource consumption at the data owner end is very low. It also privacy-preserving outsourced frequent itemset mining solution for vertically partitioned databases. This allows the data owners to outsource mining task on their joint data in a privacy-preserving manner. Based on this solution, system built a privacy preserving outsourced association rule partitioned databases. Compared with most existing solutions, this solutions leak less information about the data owners' raw data.

III. PROPOSED SYSTEM

Proposed system consists of three modules, such as user, dataset provider and third cloud. Here first query is fired by client in encrypted form is provided to cloud. Then cloud provides k closest records to the client in encrypted form. Dataset provider provides data to the cloud.

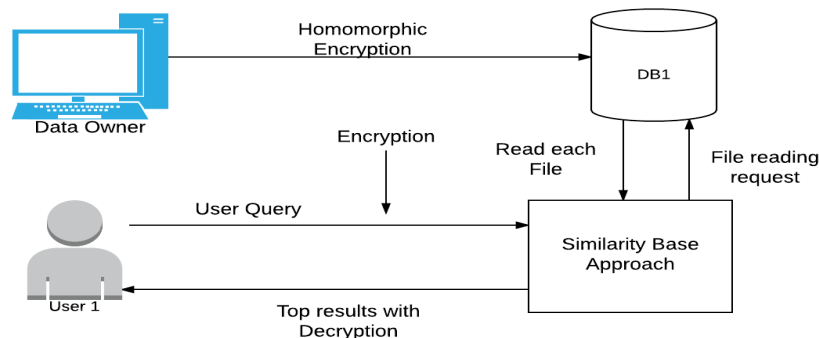


Fig.1. propose system architecture



A Secure and Dynamic Multi-keyword Ranked Search Scheme over Encrypted Cloud Data System construct a special tree-based index structure and propose a “Greedy Depth-first Search” algorithm to provide efficient multi keyword ranked search. The proposed scheme can achieve sub-linear search time and deal with the deletion and insertion of documents flexibly. Extensive experiments are conducted to demonstrate the efficiency of the proposed scheme. The typical participants of a secure search system in the cloud involve the cloud server, the data owner, and the data user, as shown in Fig.1. The data owner outsources the encrypted dataset and the corresponding secure indexes to the cloud server, where data can be encrypted using any secure encryption technique, while the secure index is generated by some particular search-enabled encryption techniques.

IV. PROPOSED ALGORITHM

1 Vector base cosine similarity (VCS)

Input: Query Q, Threshold t

Here system have to find similarity of two vectors: $\vec{a} = (a_1, a_2, a_3, \dots)$ and $\vec{b} = (b_1, b_2, b_3, \dots)$, where a_n and b_n are the components of the vector (features of the document, or values for each word of the comment) and the n is the dimension of the vectors:

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

Step 1: Read each row R from dataset D

Step 2: for each (Column c from R)

Step 3: score= Formula1(R,Q)

If(score > t)

Break;

Early stop;

Else step 2 continue

End for

Output: duplicate if returns 1 else unique

2. Mathematical Model

Let's,

Here, CS is the all system module which holds the overall system

CS={C1,C2,C3...Cn}

C1= Authentication and key generation process.

C2= upload file with data encryption and data deduplication checking.

$$\text{key}[] = \sum_{k=0}^4 (a, b, p, g)$$

Input Query Q, Threshold t show duplicate if returns 1 else unique

Here system have to find similarity of two vectors: $\vec{a} = (a_1, a_2, a_3, \dots)$ and $\vec{b} = (b_1, b_2, b_3, \dots)$, where a_n and b_n are the components of the vector (features of the document, or values for each word of the comment) and the n is the dimension of the vectors:

$$\vec{a} \cdot \vec{b} = \sum_{i=1}^n a_i b_i = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \dots (1)$$

C3= search cipher data

C4 = decryption of data and file download.

C5= analysis graphs.

Doc={D1,D2,D3.....Dn} group of documents

Query={Q1,Q2,Q3.....Qn} set of queries

DL={Doc1,Doc2.....Docn}

Here R is web base approach which handles the parallel searching, the result of query classified into n number of result pages.

V. SIMULATION RESULTS

For the system performance evaluation, calculate the matrices for accuracy. The system is implemented on java 3-tier MVC architecture framework with INTEL 3.0 GHz i5 processor and 8 GB RAM with public cloud Amazon EC2 consol. System also evaluated the computation costs of SkNNm for varying values of k, l and K. Throughout this subsection, system fix m = 6 and n = 2000. However, system observed that the running time of SkNNm grows almost linearly with n and m.

The below tables 1 shows current system evaluation outcome

Approach	Data Records	Times in Seconds
Serial input records	2000	35
	4000	68
	6000	102
	8000	132
	1000	171

Table 1: Time Required for query processing when m = 6, k = 5 and K = 512

After the complete implementation of system evaluate with different experiments. For the second experiment system focus on time complexity of cryptography algorithm. The system take use different time for data encryption as well as data decryption purpose. The below figure 3 shows the encryption and decryption time complexity.

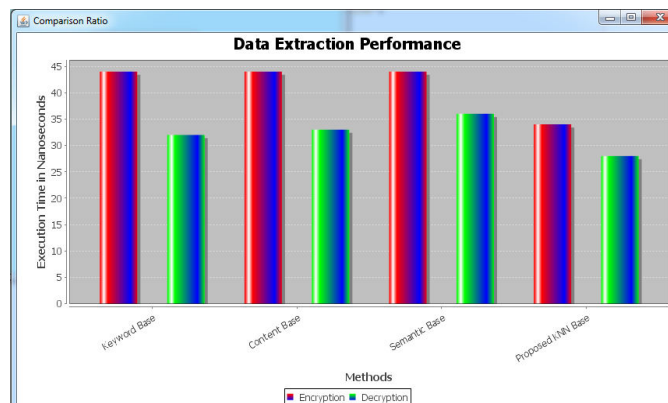


Figure 2: Data encryption and decryption performance

VI. CONCLUSION AND FUTURE WORK

To secure user privacy, numerous privacy-preserving category methods have been suggested over the past several years. The current methods are not appropriate to contracted database surroundings where the information exists in secured form on a third-party server. This paper suggested a novel privacy-preserving k-NN classification protocol over secured information in the cloud. Our protocol defends the privacy of the information, user’s input query, and conceals the information access patterns. System also analysed the efficiency of our protocol under various parameters.

For the future environment system can focus on personalize search on user feedback sessions as well as recommendation base on user point of interest with database security is the interesting part of system.

REFERENCES

1. Chi Chen et al. proposed An Efficient Privacy-Preserving Ranked Keyword Search Method IEEE 2016.
2. Chunhua Su et al. proposed Analysis and Improvement of Privacy-Preserving Frequent Item Protocol for Accountable Computation Framework IEEE 2012.
3. Cheng Huang and Rongxing Lu proposed EFPA: Efficient and Flexible Privacy-Preserving Mining of Association Rule in Cloud in IEEE 2015.
4. Bharath K. Samanthula et al. k-Nearest Neighbor Classification over Semantically Secure Encrypted Relational Data MAY 2015.
5. Lichun Li et al. Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases in AUGUST 2016.
6. Concepts and Applications of Molecular Similarity; Johnson, A.M.; Maggiora, G.M., Ed.; Wiley: New York, 1990.
7. DiMasi, J. A.; Hansen, R. W.; Grabowski, H. G. The price of innovation: new estimates of drug development costs. *J. Health Econ.* 2003, 835, 1–35.
8. Ghuloum, A.M; Sage, C.R.; Jain, A.J. Molecular hashkeys: A novel method for molecular characterization and its application for predicting important pharmaceutical properties of molecules. *J. Med. Chem.* 1999, 42, 1739 – 1748.
9. Stiefl, N.; Baumann K. Mapping Property Distributions of Molecular Surfaces: Algorithm and Evaluation of a Novel 3D Quantitative Structure-Activity Relationship Technique. *J. Chem. Inf.Comput Sci.* 2003, 46, 1390 – 1407.
10. Clark, R.D.; Cramer, R.D.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comp. Chem.* 1989, 10, 982-1012.
11. Choi, S., Ghinita, G., Lim, H.-S., & Bertino, E. (2014). Secure kNN Query Processing in Untrusted Cloud Environments. *IEEE Transactions on Knowledge and Data Engineering*, 26(11), 2818–2831. doi:10.1109/tkde.2014.2302434
12. Miss Monika D.Rokade, Mr.S.A.Kahate “processing private kNN query over untrusted data cloud”2015
13. Monika D.Rokade,S.A.Kahate, K.S..Kore “Private kNN Query Processing in Cloud Enviroments”, June 2015
14. Choi, Sunoh; Ghinita, Gabriel; Lim, Hyo-Sang; Bertino, Elisa (2014). Secure kNN Query Processing in Untrusted Cloud Environments. *IEEE Transactions on Knowledge and Data Engineering*, 26(11), 2818–2831

BIOGRAPHY

Shelke Nutan Subhash is a Research Assistant in the Computer Department, Sharadchandra Pawar College of Engineering, (Dumbarwadi), Otur, Pune. She studied Master of Engineering degree (ME) from SPPU, Pune, India. Her research interests is Cloud Computing.



INNO  SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details