SECURE AND DYNAMIC KEYWORD SEARCH TECHNIQUE WITH KNN AND LBS

Jagadeesan. D[1], Venu Priya. A[2]
[1]Professor, [2]P.G. Student, Department of Computer Science and Engineering,
Sreenivasa Institute of Technology and Management Studies, A.P., India.

ABSTRACT

The Protection of computer systems is to theft or damage to the hardware, software or the information in a secure manner. In an Existing system, SSE (Searchable Symmetric Encryption) schemes adore a high efficiency then they suffer from elaborated secret key distribution. Users have to share secret keys securely which is used for data encryption. Otherwise they are not able to share the encrypted data outsourced to the cloud. Data in Cloud is to be safe and secure manner, when a user want to retrieve LBS service oriented data from the cloud server, his location privacy is very important and also what is his requirement also to be kept securely. To achieve this proposed system "Secure and Dynamic Keyword Search Technique with kNN and LBS" uses double hashing technique and location ontology concept. For more safety Dual Servers are used, Client Server and Data Server. To safeguard the data in Data Server encryption technique is used and to get the k nearest places, kNN (k-Nearest Neighbor) algorithm is used along with location ontology concept. To maintain user's data privacy, the keyword search operation should be able to perform over encrypted data and additionally it should not leak any information about the searched keyword or the retrieved content for that double hashing technique is used. These concepts are implemented using MVC architecture and web service concept.

Keywords: Keyword search, Double hashing technique, kNN, LBS, Location Ontology.

1. INTRODUCTION

In a rapid growth development of cloud computing, user's tend to access their stored data from the remote cloud storage. The main advantage of cloud storage is ubiquitous user accessibility and also its virtually unlimited data storage capabilities. Although such benefit has been provided by the cloud, the major concern that remains is the problem over the confidentiality and privacy of data while adopting the cloud storage services [1]. For an authority, the unencrypted user data stored at the remote cloud server can be unsecured to external attacks that has been initiated by unauthorized outsiders and internal attacks launched by unreliable to Cloud Service Providers (CSPs) [2].

There are several reports that confirm data neglected related to cloud servers, due to harmful attack, embezzlement or internal errors [3]. This raises related for many users/organizations as the outsourced data might contain very sensitive personal organization/information. Many researchers have addressed the issue of ensuring confidentiality and privacy of cloud data without compromising the user functionality. Here, confidentiality refers to the secrecy of the stored data so that only the client can read the contents of the stored data. To solve the problem of confidentiality, data encryption schemes can come in handy to provide the users with some control over the secrecy of their stored data. This has been accepted by many recent researches which allows user to encrypt their data before utilizing to the cloud [4-8]. However, standard encryption schemes will dampen user's searching ability over the stored data, since after encryption a user simply cannot use a plaintext keyword to perform a search anymore and therefore cannot restore the contents in an more efficient way. The user is to search keyword that enables functionality for a certain keyword on the remote data in cloud.

Cloud application that consists of a Cloud Service Provider (CSP), and users has to store their data on cloud storage. A user's can use a traditional encryption scheme to check the confidentiality of the maintenance. A simple approach for retrieving encrypted contents related to a certain keyword, and the decrypt to perform the search on the local machine. However, this solution is not possible to do easily from a practical point of view, as the user needs to download all the contents rather than the contents containing the searched keyword. For instance, consider a scenario where the cloud storage contains 1 GB of user's data, but only 1 MB of data is related to the searched keyword. Using the simple solution, it is required to recover all the 1 GB data, which is effectively. Instead of, the user can store a plaintext keyword index in the cloud server and use it while retrieving the data. This approach will allow the CSPs to know about the keyword which is not either desirable. Accordingly, a simplicity of the data is to retrieval from a secure cloud, it need a scheme which enables user capability to search over encrypted contents [9]. A secure and efficient extraction of data is to provide, single needs to ensure that the user can perform a search over the encrypted data without striking the maintenance and the searched keyword to the server.

The cryptographic primitive that provides this characteristic is extensively known as searchable encryption (SE). This survey aims to study the searchable encryption scheme in detail and implements a solution that enables Privacy preserving data storage and extraction in Cloud computing (PrivCloud).

For our performance, it has chosen an hash coded searchable algorithm [10]. As another primary contribution, define a new version of the Smooth Projective Hash Functions (SPHFs) referred to as Linear and Homomorphic SPHF (LHSPHF). To explain the feasibility of our new framework, it is to provide an effective instantiate of the general framework from a DDH based LH-SPHF and show that it can achieve the strong security against KGA.

1.1 k-Nearest Neighbor (kNN)

k-Nearest Neighbor (kNN) is a supervised learning algorithm where the result of new method query is classified based on majority of k-Nearest Neighbor category. It is one of the most popular algorithms for pattern recognition. The purpose of this algorithm is to classify a new object based on attributes and training samples. The k-Nearest Neighbor algorithm used neighborhood classification as the prediction value of the new query method. Many analyses have found that the kNN algorithm manages a very excellent performance in their analysis on distinct data sets. In pattern recognition field, kNN is one of the most important non-parameter algorithms and it is a supervised learning algorithm. In a Euclidean distance is used as distance measured, it is only applicable to continuous variables [11].

In Figure-1 illustrate the example of kNN Classification. In a sample test (green circle) should be classified to the first class of blue squares or the second class of red triangles. If k=3 (solid line circle) it is assign to the second class because there are 2 triangles and only 1 square inside the inner circle. If k=5 (dashed line circle) it is assigned to the first class.

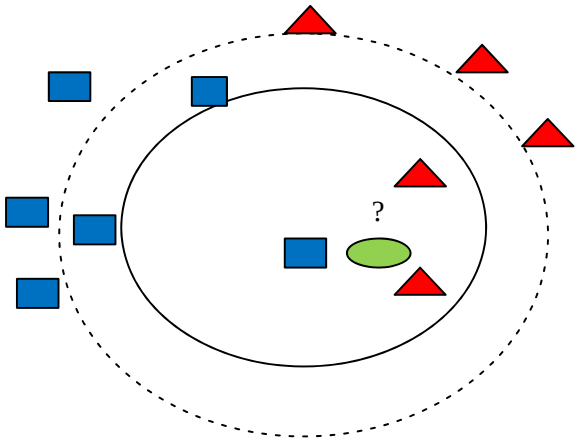

Figure -1 Example of kNN Classification

1.2 Location Based Service (LBS)

A location-based service (LBS) is an application-level service that uses location data to control aspects. LBS is an data service and has a number of uses in social networking present as information, in entertainment or security, which is available to uses information on the geographic

position of the particular location. LBS can be used in a field of contexts, such as health, indoor object search, entertainment, work, personal life, etc. LBS are very important to many businesses as well as government management to run real insight from information tied to a specific location where activities take place. LBS consist of services to identify a location of a person or thing, such as identifying the nearest banking cash machine (ATM), Hotels, Banks, Institutions, colleges, etc [12]. Geo tag includes location tag and searching query based on our location tag and searching query data will be sorted while it is displaying output.

2. RELATED WORK

D.Hyun Yum, D.S.Kim, J.S.Kim, and S.J.Hong have proposed "Order-Preserving Encryption for Non-uniformly Distributed Plaintexts", Order-preserving encryption (OPE) is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plaintexts. The BCLO scheme uses a sampling algorithm for the hyper geometric distribution as a subroutine and maps the Euclidean middle range gap to a domain gap. It is to utilize the (non-uniform) distribution of the plaintext-space to reduce the number of sampling algorithm invocations in the BCLO scheme. Instead of the Euclidean middle range gap, map the probabilistic middle range gap to a domain gap. In this it shows simulation shows that the proposed method is effective for various distributions and especially for distributions with small variance [13].

R.Curtmola, J.Garay, S.Kamara and R.Ostrovsky have proposed "Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions", Searchable symmetric encryption (SSE) allows a party to outsource the storage of his data to another party in a private manner, while maintaining the ability to selectively search over it. This problem has been the focus of active research and several security definitions and constructions have been proposed. In this paper, by reviewing existing notions of security and propose new and stronger security definitions. Then two constructions that we show secure under our new definitions. In addition to satisfying stronger security guarantees, our constructions are more efficient than all previous constructions. Further, prior work on SSE only considered the setting where only the owner of the data is capable of submitting search queries. Consider the natural extension where an arbitrary group of parties other than the owner can submit search queries. SSE in this multi-user setting, and present an efficient construction [14].

L.Fang, W.Susilo, C.Ge and J.Wang have proposed "Public key encryption with keyword search secure against keyword guessing attacks without random oracle", The notion of public key encryption with keyword search (PEKS) was put forth to enable a server to search from a collection of encrypted emails given a "trapdoor" (i.e., an encrypted keyword) provided by the receiver. The nice property in this scheme allows the server to search for a keyword, given the trapdoor. There have been subsequent works that have been proposed to enhance this notion. Two important notions include the so-called keyword guessing attack and secure channel free. The former realizes the fact that in practice, the space of the keywords used is very limited, while the latter considers the removal of secure channel between the receiver and the server to make PEKS practical. In an existing construction of PEKS secure against keyword guessing attack is only secure under the random oracle model, which does not reflect its security in the real world. Furthermore, there is no complete definition that captures secure channel free PEKS schemes that are secure against chosen keyword attack, chosen ciphertext attack, and against keyword guessing attacks, even though these notions seem to be the most practical application of PEKS primitives. In this paper, we make the following contributions. First, we define the strongest model of PEKS which is secure channel free and secure against chosen keyword attack, chosen ciphertext attack, and keyword guessing attack. In particular, we present two important security notions namely IND-SCFCKCA and IND-KGA. The former is to capture an inside adversary, while the latter is to capture an outside adversary. Second, we present a secure channel free PEKS scheme secure without random oracle under the well known assumptions [15].

A.Groce and J.Katz "A New Framework for Efficient Password-Based Authenticated Key

Exchange", Protocols for Password-based Authenticated Key Exchange (PAKE) allow two users only share a low-entropy password to agree on a cryptographically strong session key. The challenge in designing such protocols is that they must be immune to offline dictionary attacks in which an passwords in an attempt to match a password to the set of observed transcripts. A general frameworks for constructing PAKE protocols in the standard model are known. Here, we abstract and generalize a protocol to give a new methodology for realizing PAKE without random oracles, in the common reference string model. In addition to giving a new approach to the problem, the resulting construction offers several advantages over prior work. It is also an extension of our protocol that is secure within the universal composability (UC) framework and is more efficient than a previous protocol [16].

K.D.Katz and P.M.Mane "An Encrypted and Searchable Audit Log", Audit log is important part of any software system. These audit logs may contain sensitive information, poses a threat to privacy and information security, so it should be prohibited against any illegal reading and alteration or deletion. The best way is to do this as an encryption. The key challenges in an encrypted audit log are speed of log and search process, correctness of query and relevance of search results and log size. An approach for constructing searchable encrypted audit logs which can be combined with any number of existing approaches for creating tamper-resistant logs. Even though searchable encryption schemes allow users to search encrypted data by keywords securely, these techniques only support exact keyword search and will fail if there are some spelling errors are used. In this paper, the solution for fuzzy keyword search over encrypted data. K-grams are used to produce fuzzy results. Our technique for keyword search on encrypted data has wide application beyond searchable audit logs [17].

Y.Elmehdwi, B.K.Samanthula and W.Jiang "Secure kNearest Neighbor Query over Encrypted Data in Outsourced Environments", For an query processing on relational data has been studied extensively, and many theoretical and practical solutions to query processing have been proposed under various scenarios. With the recent popularity of cloud computing, users have the opportunity to outsource their data as well as the data management tasks to the cloud. Due to the rise of various privacy issues, sensitive data (e.g., medical records) need to be encrypted before outsourcing to the cloud. In query processing tasks should be handled by the cloud. Otherwise, there would be no point to outsource the data at the first place. To process queries over encrypted data without the cloud ever decrypting the data is a very challenging task. In this paper, we focus on solving the k-Nearest Neighbor (kNN) query problem over encrypted database outsourced to a cloud: a user issues an encrypted query record to the cloud, and the cloud returns the k closest records to the user. We first present a basic scheme and demonstrate that such a naive solution is not secure. To provide better security, we propose a secure kNN protocol that protects the confidentiality of the data, user's input query, and data access patterns. Also, we empirically analyze the efficiency of our protocols through various experiments. These results indicate that our secure protocol is very efficient on the user end, and this lightweight scheme allows a user to use any mobile device to perform the kNN query [18].

Y. Sun, T. La Porta, and P. Kermani "A flexible privacy enhanced location-based services system framework and practice", Location based services (LBS) are becoming increasingly important to the success and attractiveness of next generation wireless systems. However, a natural tension arises between the need for user privacy and the flexible use of location information. In this paper, a framework to support privacy enhanced location based services. It is to classify the services according to several basic criteria and we propose a hierarchical key distribution method to support these services. The main idea behind the system is to hierarchically encrypt location information under different keys, and distribute the appropriate keys only to group members with the necessary permission. Four methods are proposed to deliver hierarchical location information while maintaining privacy. It is to propose a key tree rebalancing algorithm to maintain the rekeying performance of the group key management. Furthermore, we present a practical LBS system implementation. Hierarchical location information coding offers

flexible location information access which enables a rich set of location based services. Our load tests show such a system is highly practical with good efficiency and scalability [19].

3. PROBLEM IDENTIFICATION

A Public key encryption with keyword search (PEKS) it is useful in many applications of cloudstorage. In the existing approach, the Diffie Hellman algorithm secured keyword search process consumes that enables a user to search encrypted data in the asymmetric encryption. In this paper, the authors have proposed an k-Nearest Neighbour algorithm it can provide to find both location and data privacy [20].

4. PROPOSED ARCHITECTURE

In our proposed system "Secure and Dynamic Keyword Search Technique with kNN and LBS" uses double hashing technique and location ontology concept. To safeguard the data in Data Server encryption technique is used and to get the K nearest places, kNN (k-Nearest Neighbor) algorithm is used along with location ontology concept [21]. To maintain user's data privacy, the keyword search operation should be able to perform over encrypted data and additionally it should not leak any information about the searched keyword or the retrieved content for that double hashing technique is used [22]. Geo tag includes location tag and searching query based on our location tag and searching query data will be sorted while it is displaying output [23].

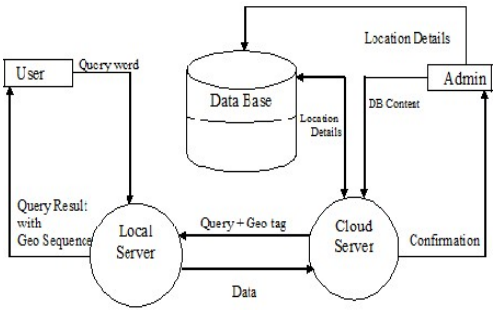

Figure-2 System Architecture of Keyword Search with locationDetails

4.1 Proposed Algorithm

k-Nearest Neighbour (kNN) Algorithm :

The steps in the Algorithm is as follows.

Step 1: Give the query keyword.

Step 2: Select all data related to that query word.

Step 3: If no results found.

Stop

Step 4: else

Retrieve all the post belongs to the keyword and in given range.

Step 5: Take K value and range

If(k > no. of posts)

Repeat step 4 with increasing range

Step 6: Calculate distance between all posts and user's (co-ordinates)

Step 7: Sort according to the distance (ascending order)

Step 8: else

Take first K results

Send result.

Step 9:Exit

4.2 Flow Diagram

As shown in the data flow diagram (Figure-3), the four activities are

Identify the location,

Enter the key word to search in that location,

Using KNN Search we have to enter k value and key word, Analyze the result.

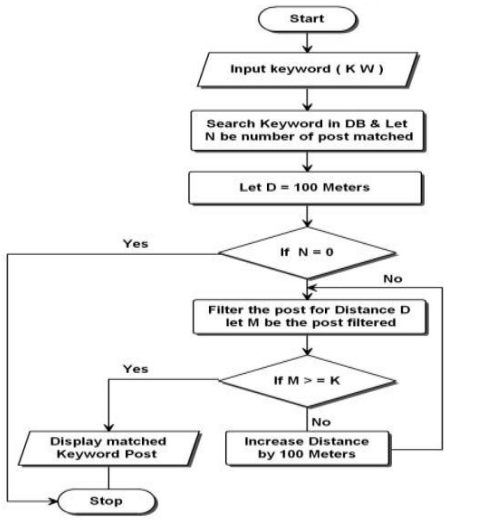

Figure-3 KNN Algorithm

5. SIMULATION RESULTS

The simulation is carried out using the double hashing technique in Java programming. The authors has been evaluated the performance of the keyword search by using the k-Nearest Neighbor algorithm [24]. The performance of the kNearest Neighbor algorithm is Hellman key algorithm [25]. The k-Nearest Neighbor algorithm is better than the Diffie Hellman algorithm is to be whatever we are entering in an postings it is displaying as an result.

6. CONCLUSION

The kNN query service has been studied based on the LBS range query service. The security of both the perturbed data and the protected queries is carefully analyzed and with encryption maintains of the security. And addresses the query privacy problem with requires the authorized query users with Dual based server process using kNN queries. Further improve the performance of query processing for both range queries and kNN queries, formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality.

References

[1] M.I.Salam et.al., "Implementation of searchable symmetric encryption for privacy-preserving keyword search on cloud storage",in Proc. HUMAN-CENTRIC COMPUTING AND INFORMATION SCIENCES, JULY 2015.

[2] R.Curtmola et.al., "Cryptographic cloud storage", in Proc. Financial Cryptography and Data Security, Springer, Berlin, Heidelberg, 2010, pp 136–149.

[3] H.Hacigumus et.al., "Executing sql over encrypted data in the database-service-provider model", in Proc. of SIGMOD, ACM, 2002,pp 216–227.

[4] S.Subashini and V.Kavitha "A survey on security issues in service delivery models of cloud computing", Journals in Networking Computer Application,2011,pp 1–11.

[5] J.Kubiatowicz et.al., "Oceanstore: an architecture for globalscale persistent storage. In: Architectural support for programming languages and operating systems", in Proc. ACM, 2000,pp 190–201.

[6] A.Muthitacharoen et.al.,"Ivy: a read/write peer-to-peer filesystem", in Proc. 5th symposium on Operating System Design and Implementation,vol. 36, 2002,pp 31–44.

[7] A.Adya et.al., "Farsite: federated, available, and reliable storage for an incompletely trusted environment", in Proc. of the 5th Symposium on Operating systems design and implementation, vol. 36, 2002,pp 1–14.

[8] J.Benaloh et.al., "Patient controlled encryption: ensuring privacy of electronic medical records", in Proc. ACM workshop on Cloud computing security, ACM, 2009,pp 103–114.

[9] M.Li, W.Lou and K.Ren "Data security and privacy in wireless body area networks", in Proc. IEEE Wireless Communications Magazine, vol. 17, IEEE, 2010,pp 51–58.

[10] M.Li et.al., "Securing personal health records in cloud computing: patient-centric and finegrained data access control in multi-owner settings", in Proc. Security and Privacy in Communication Networks, Springer, Berlin Heidelberg,2010, pp 89–106.

[11] S.Kaghyan and H.Sarukhanyan "Activity Recognition using K-Nearest Neighbour Algorithm", in Proc. Information Models and Analyses Vol.1, 2012.

[12] Y. Sun et.al., "A Flexible Privacy enhanced Location Based Services System Framework and Practice", in Proc. Networking and Security Research Center,2009.

[13] D.Hyun Yum, D.S.Kim, J.S.Kim, and S.J.Hong, "OrderPreserving Encryption for Non-uniformly Distributed Plaintexts", Proc.12thInternationalWorkshop,WISA2011,Jeju Island, Korea, August 2011, pp 22-24.

[14] R. Curtmola et.al., "Searchable SymmetricEncryption: Improved Definitions and Efficient Constructions", in Proc. 13th ACM Conference on Computer and Communications Security, 2007.

[15] L.Fang et.al., "Public Key Encryption with Keyword Search against Keyword Guessing Attacks without Random Oracle", in Proc. Information Security, 2013, pp 221-241.

[16] A.Raghavendra Praveen kumar et.al., "A Secure and Dynamic Multi Keyword Ranked Search Scheme over encrypted", in Proc. International Journal of Computer Engineering In Research Trends (IJCERT), vol 2, Issue 12, December 2015, pp 1137-1141.

[17] Komal D.Kate and P.M.Mane "An Encrypted and Searchable Audit Log", in Proc. International Journal of Engineering Research and Technology (IJERT), vol 3, Issue 5, May 2014.

[18] Y.Elmehdwi et.al., "Secure k-Nearest Neighbor Query over Encrypted Data in Outsourced Environments", in Proc. IEEE 30th International Conference on Data Engineering (ICDE), 2013.

[19] Y.Sun, T.F.La Porta and P.Kermani "A Flexible Privacy Enhanced Location Based

Services System Framework and Practice", in Proc. IEEE Transactions on Mobile Computing, 2009.

[20] D. Khader, "Public key encryption with keyword search based on K-resilient IBE," in Proc. Int. Conf. Comput. Sci. Appl. (ICCSA), 2006, pp. 298–308.

[21] J. Baek et.al., "Public key encryption with keyword search revisited," in Proc. Int. Conf. Comput. Sci. Appl. (ICCSA), 2008, pp. 1249–1259.

[22] S.Yang et.al., "A localization algorithm based on compressive sensing by K-nearest Neighbor classification", in Proc. IEEE 13th International Conference on Signal Processing (ICSP),2016.

[23] X.Lin et.al., "K Nearest Neighbors search considering traffic restriction for Location Based Service", in Proc. 17th International Conference on Geoinformatics, 2009.

[24] B.Wang et.al.,"Practical and secure nearest neighbor search on encrypted large-scale data", in Proc. 35th Annual IEEE International Conference on Computer Communications (INFOCOM), 2016.

[25] X.yi et.al., "Practical Approximate k Nearest Neighbor Queries with Location and Query Privacy", in Proc. IEEE Transactions on Knowledge and Data Engineering, vol 28, Issue 6, 2016, pp 1546-1559.