Conceptual Graph a Smart way of achieving Semantic Search over Encrypted Data Source

Kahekashan Uzma
B.Tech IV CSE Student, VITS (N6), Karimnagar, JNTUH, Hyderabad, TS, INDIA
kahekashanuzma09@gmail.com

K.Rajeevi
B.Tech IV CSE Student, VITS (N6), Karimnagar, JNTUH, Hyderabad, TS, INDIA
rajeevi.katta123@gmail.com

K.Balakishan
B.Tech IV CSE Student, VITS (N6), Karimnagar, JNTUH, Hyderabad, TS, INDIA
konukatibalakishan15@gmail.com

J.Mounika
B.Tech IV CSE Student, VITS (N6), Karimnagar, JNTUH, Hyderabad, TS, INDIA
mounikakajalli.534@gmail.com

P.Pradeep Kumar
HOD-CSE, Department of CSE, VITS, Karimnagar, JNTUH, Hyderabad, TS, INDIA
pkpuram@yahoo.com

ABSTRACT

Cloud Computing is an information technology paradigm that enables ubiquitous access to sharepool of configurable system resources and the main advantage of it is its low cost. The main drawback of Cloud Computing is its security. Many researchers have come up with a solution of encryption in cloud computing but faces a lot of difficulties.

Searchable encryption is an important area which has to be focused on Cloud Computing. The most efficient and reliable is cipher text search schemes which focuses on keyword, or shallow semantic parsing but also faces the problem of its inefficiency of satisfying user search intention. In order, to make semantic search more smart we have come up with an idea of context-aware search scheme.

In this paper we introduced conceptual graphs as knowledge representation tool. Based on the demand we have introduced two schemes as PRSCG and PRSCH-TF. To avoid the problem with the single keyword we have come up with an idea of multi keyword ranked search over encrypted cloud

data and by concentrating on PRSCG and PRSCG-TF to resolve the problem of privacy-preserving smart semantic search based on CGS. In this paper, we in deep analyzed the privacy and efficiency of proposed scheme and found it to be efficient enough.

I. INTRODUCTION

Nowadays, a large number of data owners decide to store their individual data in the cloud which can help them attain the on-demand high-quality applications and services. It also reduces the cost of data management and storage facility spending. Due to the scalability and high efficiency of cloud servers, the way for public data access is much more scalable, low-cost and stable, especially for the small enterprises. However, data owners are puzzled by the privacy of data and existing schemes prefer to use data encryption to solve the problem of information leakage. How to realize an efficient searchable encryption scheme is a challenging and essential problem. To solve the problem, we introduce Conceptual Graph (CG) as a knowledge representation tool in this paper. CG is a structure for knowledge representation based on first logic. They are natural, simple and fine-grained semantic representations to depict texts. A CG is a finite, connected and bipartite graph [1].

Our previous study [2] is able to realize the goal of performing search on CG, it's an initial and intuitive scheme which is cost expensive and not efficient.

In this paper, we propose two practical and processing schemes to solve the challenging problem CG match in the encrypted form. As a knowledge representation, CG is a perfect and mature manifestation of semantics. Since the generation of CG, it has been widely applied in many scenarios. That's why we pick up CG among various ways of knowledge representation. In order to conduct numerical calculation, we change the original CG into its linear form. When extracting CGs from original documents, we have two options according to the different aspects. One is transferring all sentences in the documents into CGs, namely PRSCG-TF. The other is sorting out the most important sentence and transferring it into a CG, namely PRSCG. In PRSCG-TF, we perform segmentation on CGs and attain their linear forms. We can view every part of the linear form for a CG as a whole. That means we can separate a CG into several individuals and regard them as "keywords" with enough semantic information. We count the TF values of these specific parts and store them in the file. Then we rank them in descending order according to TF values and select k "keywords" as representatives of the original document. Finally, we generate a dictionary to construct a numerical vector to replace the document according to vector space model. we summarize our contributions briefly as follows:

1) We use Conceptual Graphs as a knowledge representation to substitute traditional keywords and solve the Problem of privacy-preserving smart semantic search based on conceptual graphs over encrypted outsourced data. Compared with [2], it's more secure and efficient.

2)We creatively propose a modified linear form of conceptual graphs which makes quantitative calculation on conceptual graphs possible. In a sense, we facilitate fuzzy retrieval on conceptual graphs in semantic level.

3)We present two practical schemes from different aspects

to solve the problem of privacy-preserving smart seman-tic search based on conceptual graphs over encrypted outsourced data. They are both secure and efficient, but have their own focus on different aspects.

II. RELATED WORK

With the development of searchable encryption, many existing schemes provide more abundant retrieval function based on text search.

References [3]–[6] mainly discuss the single keyword search in the encrypted form. Song et al. [3] is the first to put forward the symmetric search able encryption scheme. To search over the encrypted documents with a sequential scan, the scheme employs a 2-layered encryption structure. It is the first practical scheme that defines the problem of searching on encrypted data, which has a positive effect for later researches. But its weakness is also distinct that the scheme only accepts the output of a fixed length and is suitable for its two layer encryption method and fails on variable query as well as compressed data. References [4] and [5] are proposed to make an improvement of security definition and search efficiency. An effective searchable symmetric encryption scheme is proposed in [6] to realize the ranked keyword search. The scheme uses an inverted index to store keywords and their corresponding files.

References [7]–[10] mainly focus on multiple keywords search in the encrypted form. Especially, [7] is the first one to solve the problem of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing against two threat model which is called MRSE. The paper employs vector space model and secure inner product to realize the high efficiency of search. Reference [8] generates its search index with term frequency and the vector spacemodel and chooses cosine similarity to compare the source and the query which can help achieve more accurate search results. Reference [09] provides an additional reference about how to return the ranked results through the frequency of keyword access. Reference [10] introduces parallel computing to increase the effectiveness of multi-keyword. References [12] and [14]–[15] proposes the schemes of secure outsourcing search over encrypted data. Reference [11] proposes an innovative semantic search scheme based on the concept hierarchy and the semantic relationship between concepts in the encrypted datasets. References [13] and [16]–[17] are also important studies, which study the security problems in cloud. However, the keyword still carries less semantic information. In this paper, we attempt to solve the problem of encrypted search based on CG as fast as keyword search.

III EXISTING SYSTEM:

Nowadays, a large number of data owners decide to store their individual data in the cloud which can help them attain the on-demandhigh-quality applications and services.

Many existing recent schemes are keyword-based search including single keyword and multi-keywords etc.

These schemes allow data users to retrieve interested files and return related documents in the encrypted form.

And some existing schemes hope to explore the relationships among keywords to expand the retrieval results.

So exploring a new knowledge representation with more semantic information compared with keywords to realize searchable encryption is a challenging and essential task.

IV DISADVANTAGES OF EXISTING SYSTEM:

Due to the scalability and high efficiency of cloud servers, the way for public data access is much more scalable, low-cost and stable, especially for the small enterprises.

Data owners are puzzled by the privacy of data and existing schemes prefer to using data encryption to solve the problem of information leakage.

How to realize an efficient searchable encryption scheme is a challenging and essential problem.

V PROPOSED SYSTEM:

In this paper, we propose a content-aware search scheme, which can make semantic search smarter. First, we introduce conceptual graphs (CGs) as a knowledge representation tool.

To solve the problem, we introduce Conceptual Graph (CG) as a knowledge representation tool in this paper.

CG is a structure for knowledge representation based on first logic. They are natural, simple and fine grained semantic representations to depict texts.

A CG is a finite, connected and bipartite graph.

we propose two practical and processing schemes to solve the challenging problem - CG match in the encrypted form.

As a knowledge representation, CG is a perfect and mature manifestation of semantics

VI ADVANTAGES OF PROPOSED SYSTEM:

First, we introduce conceptual graphs (CGs) as a knowledge representation tool. Then, we present our two schemes (PRSCG and PRSCG-TF) based on CGs according to different scenarios.

Second, we employ the technology of multi-keyword ranked search over encrypted cloud data as the basis against two threat models and raise PRSCG and PRSCG-TF to resolve the problem of privacy-preserving smart semantic search based on CGs.

Finally, we choose a real-world data set: CNN data set to test our scheme. We also analyze the privacy and efficiency of proposed schemes in detail. The experiment results show that our proposed schemes are efficient.

VII PROBLEM FORMULATION:

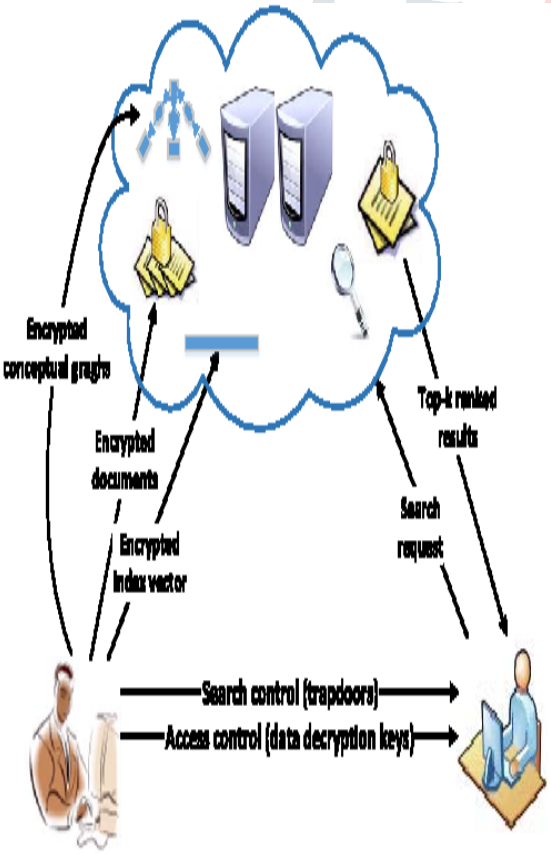


Fig. 1. The architecture of smart search based on over encrypted cloud data.

We summarize our system model showed in Fig. 1 which includes three entities: data owner, data user and cloud server.

1) Data Owner:

Data owner owns n data files

F ={F1,F2,...,Fn}that he encrypts his source documents before they are outsourced to the cloud server. Also, he must guarantee that these documents can be searched effectively. In this paper, the data owner encrypts their documents set and generates searchable indexes before outsourcing data to the cloud server. Besides this, the pre-process work such as the construction of CG, the transformation of CG into vectors and the update operation of documents should be handled ahead of time. The data user also should make a secure distribution of the key information of trapdoor generation and provide authorization for authorized data users.

2) Data Users: Data users should obtain a warrant from data owner to have access to documents. Data users should submit a simple sentence to generate a trapdoor and take back the documents which meet his requirement from the cloud server.

3) Cloud Server: Cloud server receives the store request from the data owner and execute the operation of storing the encrypted documents and searchable indexes. When the data users send the trapdoor to the cloud server, the cloud server makes a computation of relevance scores and returns top k related documents to the data users. The cloud server is also

responsible for executing the command of updating documents and searchable indexes.

Conceptual Graph: Conceptual graph as a knowledge representation model was proposed by Sowa [15]. It is defined as a graph representation for logic, which is based on the semantic networks of Artificial Intelligence (AI) and existential graphs [16]. There are usually two kinds of nodes: concepts (rectangles) and conceptual relations (ovals) (Fig. 2).

A concept is connected with another concept by conceptual relation. Each conceptual relation must be connected to some other concepts. For each CG, we named conceptual relations as semantic roles, it has 30 relations approximately include.

In this paper, we choose approximately 24 relations, regardless of tenses. Person and City inFig.2intheCGare concept types and represents the category of concepts. They can be null and marked as "*". Linear Form: Graph matching is more complex than tree matching and conceptual graph as a graph is also inconvenient for direct use. So Sowa proposes the linear form of CG which aims at solving the problem of representing multidimensional relationships in CG. It selects the node with the maximum number of edges as the vertices of tree. Then, according to the original CG, it can further form the subtree.

For Fig. 2, it can be divided into three subtrees (Fig.3) which simplify the primitive form in order of easily understanding.

In this paper, we introduce this representation with some modification.We replace concept values with concept types by leaving out concept values if there exists at the same time (Fig.4).Throug the modification, we can predigest the retrieve easily.

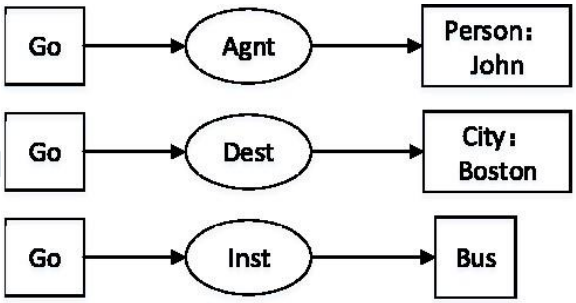
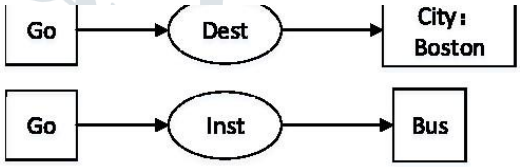


Fig 2



CG display in its linear form for *John is going to Boston*

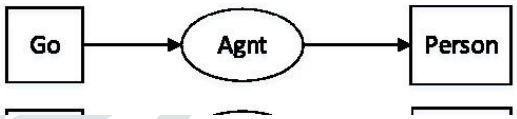


Fig 3

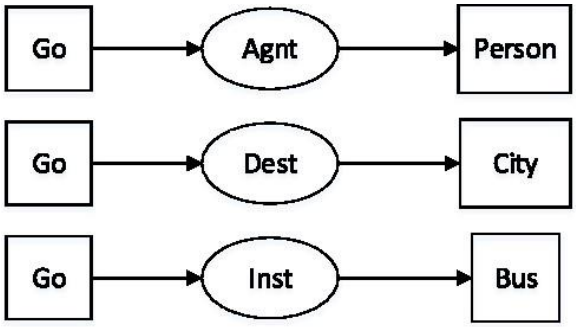


Fig 4

VIII CONCLUSION:

In this paper, compared with the previous study, we propose two more secure and efficient schemes to solve the problem of Privacy preserving smart semantic search based on conceptual graphs over encrypted outsourced data. Considering various semantic representation tools,we select Conceptual Graphs as our semantic carrier because of its excellent ability of expression and extension. To improve the accuracy of retrieval, we use Tregex simplify the key sentence and make it more generalizable. We transfer CG into its linear form with some modification creatively which makes quantitative calculation on CG and fuzzy retrieval in semantic level possible. We use different methods to generate indexes and construct two different schemes with two enhanced schemes respectively against two threat models by introducing the frame of MRSE.

We implement our scheme on the real data set to prove its effectiveness and efficiency.

For the further work, we will explore the possibility of semantic search over encrypted cloud data with natural language processing technology.

REFERENCES

[1]S.Miranda-Jiménez, A. Gelbukh, and G. Sidorov, "Summarizing conceptual graphs for automatic summarization task," in Conceptual Structures for STEM Research and Education. Berlin, Germany: Springer, 2013, pp. 245–253.

[2] Z. Fu, F. Huang, X. Sun, A. V. Vasilakos, and C. Yang, "Enabling semantic search based on conceptual graphs over encrypted outsourced data," IEEE Trans. Serv. Comput.to be published. doi: 10.1109/TSC.2016.2622697.

[3] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Secure Privacy , May 2000,pp. 44–55.

[4] Y.-C. Chang and M.Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Proc. ACNS, 2005, pp. 391–421.

[5] R. Curtmola, J. A. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions,"in Proc. ACM CCS, 2006, pp. 79 -88.

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

[7] 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.

[8] W. Sun, B. Wang, and N. Cao, "Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,"

inProc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2013, pp. 71–82.

[09] R. Li, Z. Xu, and W. Kang, "Efficient multi-keyword ranked query over encrypted data in cloud computing,"
Future Generat. Comput. Syst.,
vol. 30, pp. 179–190, Jan. 2014.

[10] Z. Fu, X. Sun, and Q. Liu, "Achieving efficient cloud search services:
Multi-keyword ranked search over encrypted cloud data supporting parallel computing,"
IEICE Trans. Commun., vols. E98–B, no. 1,pp. 190–200, 2015.

[11] Z. Fu, X. Sun, S. Ji, and G. Xie, "Towards efficient content-aware search over encrypted outsourced data in cloud," in Proc. 35th Annu. IEEE Int.
Conf. Comput. Commun. (IEEE INFOCOM), San Francisco, CA, USA,Apr. 2016, pp. 1–9, doi: 10.1109/INFOCOM.2016.7524606.

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

[13] Y. Ren, J. Shen, J. Wang, J. Han, and S. Lee, "Mutual verifiable provabledata auditing in public cloud storage,"
J. Internet Technol., vol. 16, no.2,pp. 317–323, 2015.

[14] J. Wang, X. Chen, X. Huang, I. You, and Y. Xiang, "Verifiable auditing for outsourced database in cloud computing,"IEEE Trans. Comput.,vol. 64, no. 11, pp. 3293–3303, Nov. 2015.

[15] X. Chen, J. Li, J. Ma, Q. Tang, and W. Lou, "New algorithms for secure outsourcing of modular exponentiations,"
IEEE Trans. Parallel Distrib.Syst.
, vol. 25, no. 9, pp. 2386–2396, Sep. 2014.

[16] C. Wang, N. Cao, and K. Ren, "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.

[17] Z. Wu, B. Liang, and L. You, "High dimension space projection-based biometric encryption for fingerprint with fuzzy minutia,"Soft Comput.,vol. 20, no. 12, pp. 4907–4918, 2016.