Asian Journal of Information Technology 16 (6): 392-398, 2017
ISSN: 1682-3915

Secure Hybrid Search in Encrypted Cloud Data

[1]D. Palanivel Rajan, [1]S. John Alexis and [2]S. Pravinth Raja
[1]Coimbatore Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India
[2]Sri Ramakrishna Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract: Cloud computing grow to be ubiquitous which makes more insightful information are being centralized into the cloud. To protection the data privacy usually sensitive data have to be encrypted before outsourcing. This makes very challenging task as effective data utilization. To store data on data storage servers such as mail servers and files servers in encrypted form to reduce security and privacy risks. But this usually implies that one has to sacrifice functionality for security. For example, if a client wishes to retrieve only documents containing certain words, it was not previously known how to let the data storage server perform the search and answer the query without loss of data confidentiality. To overcome these issues traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. In this study, we define and solve the problem of effective yet, secure keyword search over encrypted cloud data. This hybrid secure search greatly enhances system usability by returning the matching files in a ranked and categorized regarding to certain relevance criteria (e.g., keyword frequency) with light weight many to many authentication along with symbol-based tree traverse search scheme will make one step closer towards practical deployment of privacy-preserving data hosting services in cloud computing.

Key words: Category based keyword search, confidential data, searchable encryption, cloud computing, hybrid search algorithm, hosting, practical

INTRODUCTION

Cloud computing enables cloud customers to store their data into the cloud and provides an on-demand high quality applications and services from a shared pool of configurable computing resources (Buyya *et al*., 2008). The benefits brought by this new computing model include but are not limited to: relief of the burden for storage management, universal data access with independent geographical locations and avoidance of capital expenditure on hardware, software and personnel maintenances, etc. With the prevalence of cloud services, more and more sensitive information are being centralized into the cloud servers such as e-Mails, personal health records, private videos and photos, company finance data, government documents, etc. (Chang and Mitzenmacher, 2005). To protect data privacy and combat unsolicited accesses, sensitive data has to be encrypted before outsourcing (Mell and Grance, 2010), so as to provide end-to-end data confidentiality assurance in the cloud and beyond.

However, data encryption makes effective data utilization a very challenging task given that there could be a large amount of outsourced data files. Besides, in cloud computing, data owners may share their outsourced data with a large number of users who might want to only retrieve certain specific data files they are interested in during a given session. One of the most popular ways to do so is through key word-based search. Such key word search technique allows users to selectively retrieve files of interest and has been widely applied in plaintext search scenarios (Chase and Kamara, 2010). Unfortunately, data encryption which restricts user's ability to perform key word search and further demands the protection of key word privacy, makes the traditional plaintext search methods fail for encrypted cloud data.

Although, traditional searchable encryption schemes (Li *et al*., 2010; Deepa *et al*., 2012; Hu and Liu, 2013) allow a user to securely search over encrypted data through key words without first decrypting it, these techniques support only conventional Boolean key word search without capturing any relevance of the files in the search result. When directly applied in large collaborative data outsourcing cloud environment, they may suffer from the following two main drawbacks. On the one hand for each search request, users without pre-knowledge of the encrypted cloud data have to go through every retrieved files in order to find ones most matching their interest which demands possibly large amount of post processing overhead; on the other hand invariably sending back all

Corresponding Author: D. Palanivel Rajan, Coimbatore Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

files solely based on presence/absence of the key word further incurs large unnecessary network traffic which is absolutely undesirable in today's pay-as-you-use cloud paradigm. In short, lacking of effective mechanisms to ensure the files retrieval accuracy is a significant drawback of existing searchable encryption schemes in the context of cloud computing. Nonetheless, the state-of-the-art in information retrieval community has already been utilizing various scoring mechanisms (Wang *et al*., 2010a, b).

To protect data privacy, sensitive data has to be encrypted before outsourcing so as to provide end-to-end data confidentiality assurance in the cloud and beyond. Thus, exploring privacy-preserving and effective search service over encrypted cloud data is of paramount importance (Song *et al*., 2000). Data owners may share their data with large number of on demand data users and huge amount of outsourced data documents in cloud, this problem is particularly challenging as it is extremely difficult to meet also the requirements of performance, system usability and scalability.

On the one hand, to meet the effective data retrieval need, large amount of documents demand cloud server to perform result relevance ranking instead of returning undifferentiated result (Petrov *et al*., 2013). Such ranked search system enables data users to find the most relevant information quickly. Category based search can also elegantly eliminate unnecessary network traffic by sending back only the most relevant data. The other hand, to improve search result accuracy as well as enhance user searching experience, it is also crucial for such ranking system to support multiple keywords search as single keyword search often yields far too coarse result. Some recent designs have been proposed to support Boolean keyword search as an attempt to enrich the search flexibility, they are still not adequate to provide users with acceptable result ranking functionality and solves the secure ranked search over encrypted data with support of only single keyword query.

Motivation: With the advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to the commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data have to be encrypted before outsourcing which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in the cloud, it is necessary to allow multiple keywords in the search request and return documents in the order of their relevance to these keywords. That is there is no tolerance of minor typos and format inconsistencies which on the other hand are typical user searching behavior and happen very frequently. This significant drawback makes existing techniques unsuitable in cloud computing as it greatly affects system usability, rendering user searching experiences very frustrating and system efficacy very low. These things create the problem in both the encryption side as well as in searchable keywords side.

Literature review: The evolution of cloud computing is one of the major advances in the technologies which represent cloud computing: Platform-as-a-Service (PaaS), Software-as-a-Service (SaaS) and Infrastructure as-a-Service (IaaS). Public key encryption (Petrov *et al*., 2013) deals with the privacy of database data. There are two different scenarios: public databases and private databases.

Private databases: A user wishes to upload its private data to a remote database and wishes to keep the data private from the remote database administrator. An additional privacy requirement is to hide any information from the database administrator regarding the access pattern, i.e., if some item was retrieved more than once, some item was not retrieved.

Public databases: The database data is public (such as stock quotes) but the user is unaware of it and wishes to retrieve some data-item or search for some data-item without revealing to the database administrator which item it is. All keyword searches are based on this index; hence our scheme does not order full pattern-matching generality with the actual text. In practice, this should be sufficient for most users. It is worth noting that this framework can have complete control over what words are keywords that can be useful for many applications. In confidentiality-preserving rank-ordered search when an authorized user remotely accesses the data to search and retrieve desired documents, the large size of the collections.

Traditional searchable encryption (Wang *et al*., 2010a, b; Matwyshyn *et al*., 2010; Venugopal *et al*., 2008; Chang and Mitzenmacher, 2005; Chase and Kamara, 2010; Mell and Grance, 2010; Li *et al*., 2010; Deepa *et al*., 2012) has been widely studied in the context of cryptography. Among those works, most are focused on efficiency improvements and security definition formalizations. The first construction of searchable encryption was proposed by Song *et al*. (2000) in which each word in the document is encrypted independently under a special two layered encryption construction. Goh (Venugopal *et al*., 2008)

proposed to use Bloom filters to construct the indexes for the data files. For each file, a Bloom filter containing trapdoors of all unique words is built up and stored on theserver. To search for a word, the user generates the search request by computing the trapdoor of the word and sends it to the server. Upon receiving the request, the server tests if any Bloom filter contains the trapdoor of the query word and returns the corresponding file identifiers. To achieve more efficient search (Chang and Mitzenmacher, 2005; Mell and Grance, 2010; Curtmola *et al*., 2006) both proposed similar\index approaches where a single encrypted hash table index is built for the entire file collection. In the index table, each entry consists of the trapdoor of a keyword and an encrypted set of file identifiers whose corresponding data files contain the keyword. As a complementary approach Chang and Mitzenmacher (2005) presented a public-key based searchable encryption scheme with an analogous scenario to that by Matwyshyn *et al*. (2010). In their construction, anyone with the public key can write to the data stored on the server but only authorized users with the private key can search. As an attempt to enrich query predicates, conjunctive keyword search, subset query and range query over encrypted data have also been proposed by Li *et al*. (2010), Deepa *et al*. (2012) and Hu and Liu (2013). Note that all these existing schemes support only exact keyword search and thus are not suitable for cloud computing.

Private matching (Narudkar and Aparna, 2015) as another related notion has been studied mostly in the context of secure multiparty computation to let different parties compute some function of their own data collaboratively without revealing their data to the others. These functions could be intersection or approximate private matching of two sets, etc., (Fu *et al*., 2014). Private information retrieval (Cao *et al*., 2011) is an often-used technique to retrieve the matching items secretly which has been widely applied in information retrieval from database and usually incurs unexpectedly computation complexity.

System architecture: Data can be stored both as public as well as private. Different searching strategies are available for both types of data. The confidential data are stored in the cloud using encryption technique (Buyya *et al*., 2008; Liu, 2010). So, only the authenticated members who know the key can access the data. Accessing data from encrypted storage is very difficult. A different type of searching technique is used to search for encrypted data (Shetty *et al*., 2012).

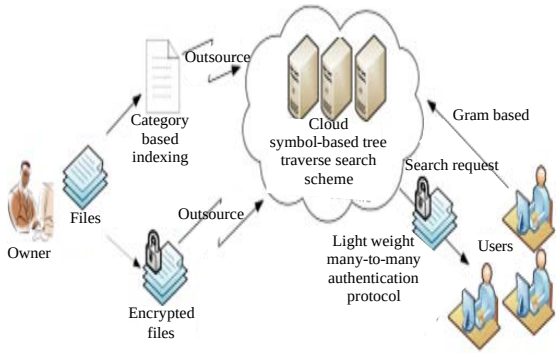

Fig. 1: Proposed secure hybrid search system in encrypted cloud platform

Figure 1 explains the different functions of the proposed secure hybrid search system in the encrypted cloud data. It consists of the four different parts as discussed.

Category based indexing: Category based indexing techniques improve accuracy and search efficiency. These techniques consolidate the keywords to obtain a better cooperative indexing solution. In this category, collaborative machine learning and collaborative knowledge representation and reasoning methods are included.

Collaborative filtering algorithm 1: Collaborative filtering based on video reindexing to detect similarity among mobile users and to predict the missing preferences. It is a concept-based filtering that refines the indexing scores of concept classifiers by exploiting structures embedded within the score matrix. The detection rate is high because various indexing algorithm types are utilized. In addition, the method is independent of classifier type and can be applied to any classification results without having to retrain the models (Algorithm 1).

Algorithm 1; File uploading and keyword extraction algorithm:

For all the files
Scan the file f_i
While new distinct word then
add the distinct words in wrd where wrd = (dw1, dw_2, …, dw_n)
End while
For each w_i in wrd
compute the score f_{ij}= $f(w_i)$
insert in the category list f_{ij}
Each word
Move to the cloud server
Encrypt file
move encrypted files to the cloud sever
End files

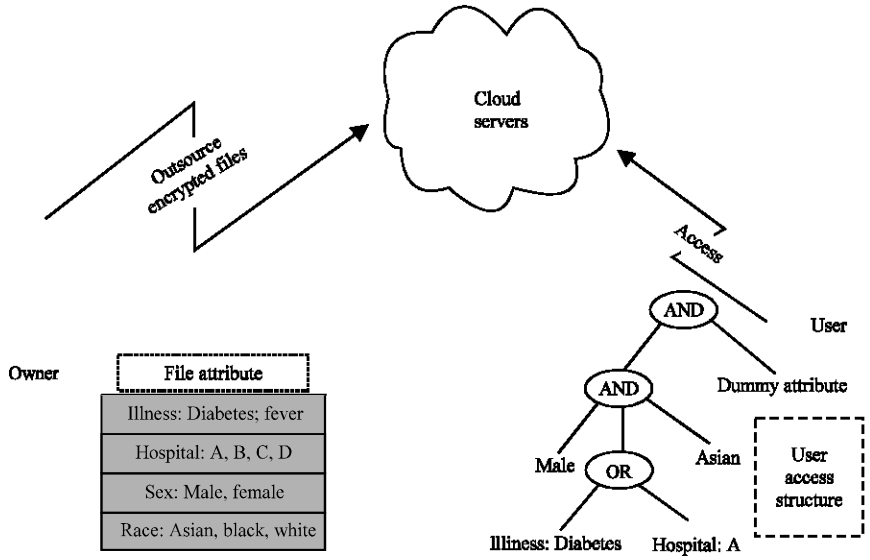

Fig. 2: Symbol based tree traverse search scheme

Symbol based tree traverse search scheme: In this technique, a multi-way tree is constructed for storing the fuzzy keyword set over a finite symbol set. All the trapdoors sharing a common prefix have common nodes. The fuzzy keyword in the trie can be found by depth first search approach (Chang and Mitzenmacher, 2005). The main advantages of this is maintaining keyword privacy and effective utilization of remotely stored which is explained in Fig. 2.

Gram based search request: The gram of a string is a substring and can be used for effective approximate search. The order of the characters after the primitive operation is always kept the same before the operations (Algorithm 2).

Algorithm 2; Gram based search request:

```
procedure create gram request (wi, d)
 if d>1 then
Call create gram fuzzy set (wi, d - 1)
 end if
 if d = 0 then
S ' wi, d = {wi}
 else
for (k-1 to |S' wi, d-1 |) do
   for j←1 to 2 * |S' wi, d-1 [k]|+1 do
     Set fuzzyword as S' wi, d-1 [k]
     Delete the j-th character
     if fuzzy word is not in S' wi, d-1 then
Set S' wi,d = S' wi, d←{fuzzyword}
     end if
end for
     end for
end if
end procedure
```

Light weight many-to-many authentication protocol: The lightweight many-to-many authentication protocol that uses near field communications as a carrier technology is proposed. The solution works without any.

User interaction and can be applied for almost any data storage device: NFC or RFID tag, USB flash drive, etc. The major novelty of the system is real-time encryption key generation algorithm. This approach doesn't require any computation power on the tag, trusted third parties or secure link between tag and information system. So far, the mentioned features transforms to signficant advantages to the cloud also while compared to other existing approaches for user authentication such as OAuth (Wang *et al*., 2010a, b), Opacity (Cao *et al*., 2011, 2014) and LMAP (Li *et al*., 2010) has to be presented. Also, the current version of the protocol allows user to modify the key sequence for any IS in an uncontrolled manner. So far, the solution in present form could be applied for user authentication.

MATERIALS AND METHODS

Figure 3 shows the steps and the interaction between the data owner, user and the cloud server is illustrated as follows. In the stage the data owner creates a category based search index for each document. The index is encrypted using the secret keys and the trapdoor is created and is shared between the data owner and the authorized users using the light weight many-to-many authentication protocol. The Data owner

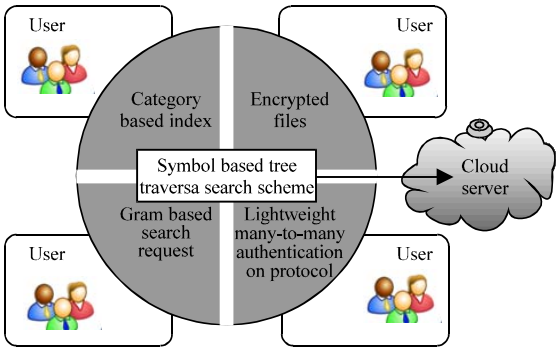

Fig. 3: Proposed secure hybrid search block diagram

encrypts the files, then uploads these search index and the encrypted files to the data server. The cloud server convert that files and search index as symbol which is helpful for searching because the server search using the symbol based tree traverse search scheme which makes the server to find the required file fast. This process is called the index generation and the secure search as shown in Fig. 1.

Step 1: When a user wants to perform a keyword search, he first connects to the encrypted server using the light weight many-to-many authentication protocol.

Step 2: Data user required to search the content available in the encrypted cloud server he submitted a search keywords and these words are converted to gram based search request in cloud server.

Step 3: Receive the gram based search query and convert that into the symbol based query.

Step 4: Encrypted server performs the symbol based tree traverse to find the required document.

Step 5: Then, the user request the decryption keys from the data owner using the light weight many-to-many authentication get the decryption keys for the files and decrypt the required files.

RESULTS AND DISCUSSION

In this study, we demonstrate a thorough experimental evaluation of the proposed Secure Hybrid Search (SHS) over Similarity Based Multiple Keywords Search (SBMKS) technique on a test data set. We randomly select different number of documents to build data set to demonstrate performance analysis.

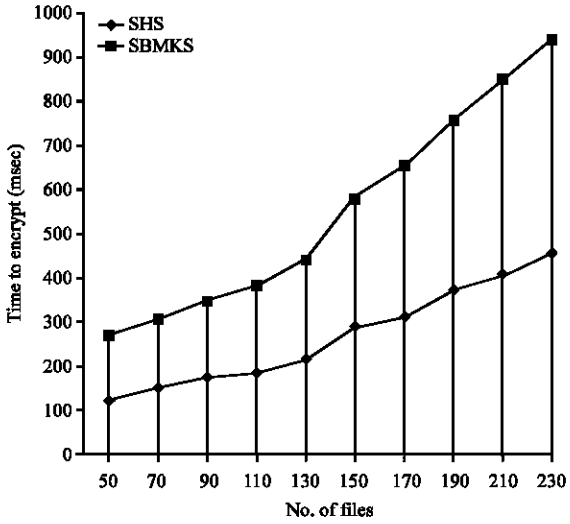

Fig. 4: Time taken for encryption

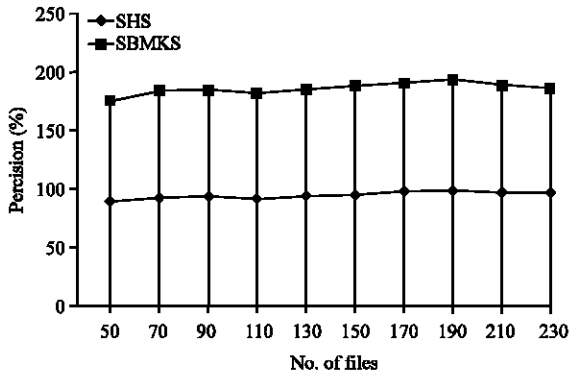

Fig. 5: Precision of the search result

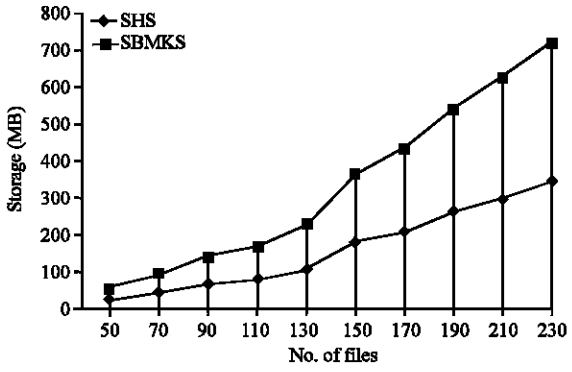

Fig. 6: Keyword storage

The result value is tabulated in Table 1-4 and Fig. 4-7 illustrate the various performance. Entire experiment system is implemented using cloud sim and java language on a Linux Server. Performance of our technique is evaluated regarding the efficiency of

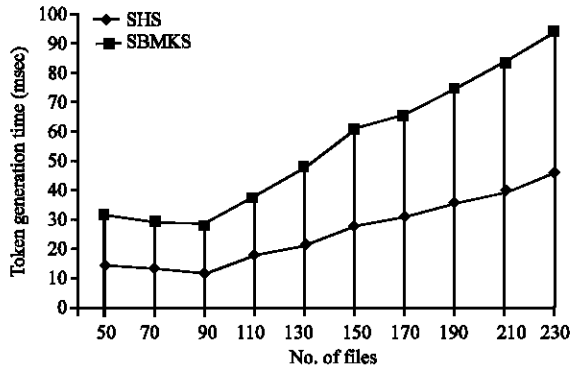

Fig. 7: Security token generation time

Table 1: Token generation time (msec)

The no. of documents	SHS	SBMKS
50	15	17
70	14	16
90	12	17
110	18	20
130	21	27
150	28	33
170	31	35
190	36	39
210	40	44
230	46	49

Table 2: Time to encrypt (msec)

The no. of documents	SHS	SBMKS
50	125	145
70	148	164
90	172	187
110	185	201
130	210	237
150	287	301
170	316	345
190	376	391
210	410	448
230	461	489

Table 3: Percision (%)

The no. of documents	SHS	SBMKS
50	91.23	86.26
70	93.45	91.56
90	94.23	92.89
110	93.68	89.96
130	94.12	92.12
150	96.16	94.23
170	97.12	95.45
190	98.89	96.16
210	96.17	94.87
230	95.32	92.12

Table 4: Storage (mb)

The No. of documents	SHS	SBMKS
50	13	33
70	36	52
90	60	75
110	73	89
130	98	125
150	175	189
170	204	233
190	264	279
210	298	336
230	349	377

traditional searching technique in cloud platform as well as the tradeoff between search precision and privacy.

CONCLUSION

In this study, the secure hybrid searching in encrypted cloud data is addressed. We formalized and solved the problem of supporting efficient yet privacy-preserving hybrid search for achieving effective utilization of remotely stored encrypted data in cloud computing. We utilized the gram-based technique to construct the storage efficient keyword sets by exploiting the similarity metric of edit distance. Based on the constructed keyword sets, we further employees a symbol-based tree traverse searching scheme where a multi-way tree structure is built up using symbols transformed from the resulted keyword sets. light weight many-to-may encryption provides the fast and secure user key encryption which makes that our proposed solution is secure and privacy-preserving with fast operations.

SUGGESTIONS

There are possible improvements and undergoing efforts that will appear in the future research. Firstly, applying the soft computing techniques and compare the performance on the keyword indexing along with proposed framework and secondly, the proposed method will be tested on a real dataset in order to compare the performance of our secure hybrid search method with the other methods used in plain datasets that do not involve any security or privacy-preserving techniques.

REFERENCES

Buyya, R., C.S. Yeo and S. Venugopal, 2008. Market-Oriented cloud computing: Vision, Hype and reality for delivering IT services as computing utilities. Proceedings of the 10th IEEE International Conference on High Performance Computing and Communications, September 26-28, 2008, Houston, USA., pp: 5-13.

Cao, N., C. Wang, M. Li, K. Ren and W. Lou, 2011. Privacy-preserving multi-keyword ranked search over encrypted cloud data. Proceedings of the 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, April 10-15, 2011, Shanghai, China, pp: 829-837.

Cao, N., C. Wang, M. Li, K. Ren and W. Lou, 2014. Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE. Transac. parallel Distrib. Syst., 25: 222-233.

Chang, Y.C. and M. Mitzenmacher, 2005. Privacy preserving keyword searches on remote encrypted data. Proceedings of the 3rd International Conference on Applied Cryptography and Network Security, June 7-10, 2005, New York, USA., pp: 442-455.

Chase, M. and S. Kamara, 2010. Structured encryption and controlled disclosure. Proceedings of the 16th International Conference on the Theory and Application of Cryptology and Information Security, December 5-9, 2010, Springer, Berlin, Germany, pp: 577-594.

Curtmola, R., J.A. Garay, S. Kamara and R. Ostrovsky, 2006. Searchable symmetric encryption: Improved definitions and efficient constructions. Proceedings of the 13th ACM Conference on Computer and Communications Security, October 30-November 3, 2006, Alexandria, USA., pp: 79 -88.

Deepa, P.L., S.V. Kumar and D.S. Karthik, 2012. Searching techniques in encrypted cloud data. Intl. J. Adv. Res. Comput. Eng. Technol. IJARCET., 1: 1-5.

Fu, Z., X. Sun, N. Linge and L. Zhou, 2014. Achieving effective cloud search services: Multi-keyword ranked search over encrypted cloud data supporting synonym query. IEEE. Trans. Consum. Electron., 60: 164-172.

Hu, C. and P. Liu, 2013. Public key encryption with ranked multi-keyword search. Proceedings of the 2013 5th International Conference on Intelligent Networking and Collaborative Systems (INCoS), September 9-11, 2013, IEEE, Xi'an, China, ISBN:978-0-7695-4988-0, pp: 109-113.

Li, J., Q. Wang, C. Wang, N. Cao, K. Ren and W. Lou, 2010. Fuzzy keyword search over encrypted data in cloud computing. Proceedings of the 9th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies, March 15-19, 2010, San Diego, CA., USA., pp: 1-5.

Liu, H., 2010. A n ew form of DOS attack in a cloud and its avoidance mechanism. Proceedings of the 2010 ACM Workshop on Cloud Computing Security Workshop, October 08-08, 2010, ACM, Chicago, Illinois, USA., ISBN:978-1-4503-0089-6, pp: 65-76.

Matwyshyn, A.M., A. Cui, A.D. Keromytis and S.J. Stolfo, 2010. Ethics in security vulnerability research. IEEE. Secur. Privacy, 8: 67-72.

Mell, P. and T. Grance, 2010. Draft NIST working de?nition of cloud com-puttin. National Institute of Standards and Technology, Gaithersburg, Maryland, USA. http://csrc.nist.gov/groups/SNS/cloud-computing/index.html.

Narudkar, A. and A.J. Aparna, 2015. Survey on searching techniques over encrypted data. IJCSIT. Intl. J. Comput. Sci. Inf. Technol., 6: 1007-1010.

Petrov, V., M. Komar and Y. Koucheryavy, 2013. A lightweight many-to-many authentication protocol for near field communications. Proceedings of the 2013 21st IEEE International Conference on Network Protocols (ICNP), October 7-10, 2013, IEEE, Goettingen, Germany, ISBN:978-1-4799-1270-4, pp: 1-2.

Shetty, S., N. Luna and K. Xiong, 2012. Assessing network path vulnerabilities for secure cloud computing. Proceedings of the 2012 IEEE International Conference on Communications (ICC), June 10-15, 2012, IEEE, Ottawa, Canada, ISBN:978-1-4577-2052-9, pp: 5548-5552.

Song, D., D. Wagner and A. Perrig, 2000. Practical techniques for searches on encrypted data. Proceeding of the IEEE Symposium on Security and Privacy, May 14-17, 2000, Berkeley, CA., USA., pp: 44-55.

Venugopal, S., X. Chu and R. Buyya, 2008. A negotiation mechanism for advance resource reservations using the alternate offers protocol. Proceedings of the 16th International Workshop on Quality of Service IWQoS, June 2-4, 2008, IEEE, Enschede, Netherlands, pp: 40-49.

Wang, C., N. Cao, J. Li, K. Ren and W. Lou, 2010b. Secure ranked keyword search over encrypted cloud data. Proceedings of the IEEE 30th International Conference on Distributed Computing Systems, June 21-25, 2010, Genoa, Italy, pp: 253-262.

Wang, J.A., H. Wang, M. Guo, L. Zhou and J. Camargo, 2010a. Ranking attacks based on vulnerability analysis. Proceedings of the 2010 43rd Hawaii International Conference on System Sciences (HICSS), January 5-8, 2010, IEEE, Honolulu, Hawaii, USA., ISBN:978-1-4244-5509-6, pp: 1-10.