



WWJMRD 2017; 3(12): 151-156
www.wwjmr.com
International Journal
Peer Reviewed Journal
Refereed Journal
Indexed Journal
UGC Approved Journal
Impact Factor MJIF: 4.25
e-ISSN: 2454-6615

Gopichand G

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Santhi H

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Gayathri P

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Shaik Naseera

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Geraldine Bessie Amali

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Correspondence:

Gopichand G

School of Computer Science
and Engineering, Vellore
Institute of Technology,
Vellore, Tamil Nadu, India

Improved Multi-Keyword Ranked Search Mechanism over Encrypted Cloud Data

Gopichand G, Santhi H, Gayathri P, Shaik Naseera, Geraldine Bessie Amali

Abstract

Because of the expanding prominence of CLOUD computing, an ever-increasing number of information proprietors are inspired for outsourcing their information to the servers of cloud for extraordinary convenience as well as lessened price in management of Data. However, encryption of the delicate data must be done before outsourcing it for the security necessities, which obsoletes utilization of data like retrieving the documents based on the keyword. In this paper, we present a protected multi-keyword ranked search algorithm over the data stored on the CLOUD server in the encoded form, where operations which are dynamic are supported by it simultaneously like deletion and insertion of any document. For constructing an index and generating the query, the vector space model and the broadly used TF-IDF model are combined. Further, a tree based index structure has been constructed and to give a productive "multi-keyword" ranked search, we proposed a "Greedy Depth-first Search" algorithm. To encode a query vector and an index, we used a secure kNN algorithm as well as the accurate relevance score calculation between query vectors and an index is ensured. With a specific end goal to resist statistics attacks, phantom terms are added to the vector of indices for blinding items in the query. Sub-linear search time can be achieved by the proposed algorithm as well as insertion and deletion are managed to ensure the flexibility.

Keywords: CLOUD Computing, Outsourcing, Multi-Keyword ranked algorithm, kNN, Encryption

Introduction

A new model of the IT foundation enterprise is known as CLOUD computing, where colossal resource of applications, storage and computing can be organized as well as clients are empowered to take pleasure of convenient, ordinary and on-demand access of network to a shared pool of configurable computing assets with incredible effectiveness and insignificant financial overhead. Both individuals and enterprises are attracted by these engaging components of the CLOUD computing. So rather than buying hardware and the software to deal with the data themselves, they are inspired to outsource their delicate data to the CLOUD. Even though there are various advantages of outsourcing the data to the CLOUD, protection concerns are brought when outsourcing the sensitive data to the remote servers like finance data of the company, id proof documents, e-mails etc. Data of the clients are kept with the CLOUD service provides who may access the delicate data without approval. So, a general way is used to encode the data before outsourcing it. But this will bring about an immense cost as far as data ease of use. For retrieving the information on the basis of keywords there exists many keywords based techniques which are mostly utilized for the plain text data. But these techniques cannot be straightforwardly applied to the encoded data. Downloading the delicate data from the cloud and then decode the same locally is clearly unfeasible. For addressing this issue, some general-purpose solutions have been outlined by the researchers with completely homomorphic encryption as well as with the oblivious RAMs. However, these strategies are impractical because of their computational overhead which is very high for both the users of cloud and the cloud server.

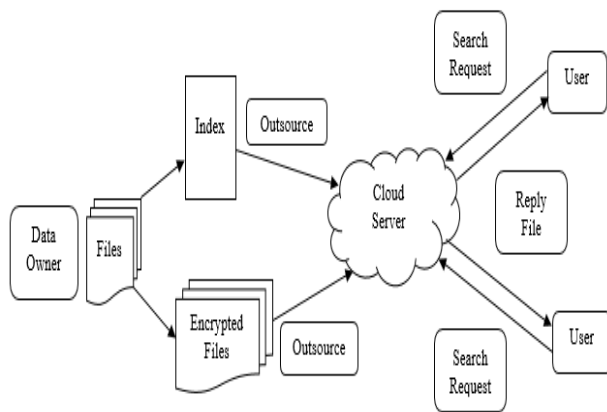


Fig. 1: Data Storage in Cloud

In fig 1, it is shown that first the data which are stored into the files are encrypted as well as indexed and then store it on the Cloud servers. When any user needs to download any data file, he sends a request to the Cloud servers for accessing the data. Data owners give an access for the downloading of the file if he is an authorized user. And then user searches with the 'keyword' and he gets all the documents which has the searched keyword present in the documents (files) with the probabilities of occurrences of the keywords in the document. And finally, the file has been sent to the user.

In the current years, numerous cipher text search algorithms have been proposed by incorporating the cryptography techniques by the researchers. These techniques have been demonstrated with provable security, but enormous operations are required by their techniques as well as time complexity is very high. In this manner, previous strategies are not appropriate for the circumstance of big data where volume of data is huge and online data dealing is required by the applications.

Also, the relationship between documents is hidden in the above strategies. The properties of the documents are represented by the relationship between the documents and henceforth keeping up the relationship is indispensable for document express completely. On the other hand, damaged data may be contained in the data search results which are coming back to the users due to hardware/software failure or corruption in the storage or intruder or the malicious administrator may distort it. In this manner, a verifiable scheme ought to be given to users to check the rightness furthermore, completeness of the search items.

In actuality, particular contributions have been made by searchable encoded (SE) schemes which are the more practical uncommon purpose solutions as far as productivity, security and functionality are considered. The customer is enabled for storing the data in encoded format to the cloud as well as searching for keywords is executed over the domain cipher text by searchable encoded scheme. Up until now, to accomplish various search functionality like multi keyword ranked search, single keyword search, ranked search, similarity search etc. under various threat model, it has proposed an abundant work. Among all these techniques, more attention has been achieved by multi-keyword ranked search for its reasonable applicability.

Recently it has proposed some dynamic schemes to bolster insertion and deletion operations for the document collections. These are critical acts as it is exceptionally

conceivable that the owners of the data need to modernize their data on the cloud server. In any case, effective multi-keyword ranked search is supported by the few of the dynamic schemes.

In this paper, we utilize the vector space model and a vector represent each and every document. Which implies each and every document can be viewed as a point in a high dimensional space. Because of the connection between various documents, every one of the documents can be separated into a few classifications. We can say that short distance points in the high dimensional space can be characterized into a particular category. The hunt time can be to a great extent lessened by choosing the coveted category and surrendering the irrelevant classifications. Contrasting and every one of the documents in the dataset, the quantity of document which user goes for is little. Because of the modest number of the coveted documents, a particular classification can be further isolated into a few sub-categories.

Rather than utilizing the conventional sequence search strategy, it creates a backtracking scheme for searching the final records. The categories will be first searched by the cloud servers as well as get the base craved sub-category. After this a desired k documents are chosen by the cloud server from the base desired sub-category. The user chooses the value of k previously and sends the same to the cloud server. On the off chance that present sub-classification cannot fulfil the k records, cloud server will follow back to its parent and select the coveted documents from its sibling categories. This procedure will be executed recursively until the coveted k documents are fulfilled or the root is come to.

To check the trustworthiness of the result, we have developed a hash function which has a verifiable structure. Each report will be hashed and for document representation, hash results will be utilized. The hashed aftereffects of records will be hashed again with the category data that these records have a place with and the outcome will be utilized to represent to the present category. Essentially, hash result of the combination of category which is used currently and the sub-categories information represent each and every category. To represent all the categories and the information associated with the same, a virtual root has been built. The virtual root is indicated by the hash result of the link of the considerable number of categories situated in the first level. The virtual root will be marked with the goal that it is verifiable. The virtual root is confirmed by the user for verifying any document rather than checking each and every report.

Literature Survey

We have studied the issue of encrypting the public key with conjunctive search of keywords (PECK). A client is empowered by the encryption of keyword search for outsourcing his sensitive data to the capacity of server which is untrusted along with the assurance that to specifically search the data which he has outsourced without any leaking in it. [5] The document search is provided by the PECK which contains each of few keywords over a public key setting. Here, initially an effective PECK algorithm has been developed which proves security over a Diffie-Hellman algorithm which is a linear and decisional assumption in the irregular model of

oracle. As compared to the past algorithms, it has the shortest cipher text size as well as an equivalent computational overhead is required. Another scheme has been introduced for overcoming the security issues as multi-user PECK scheme. [7] An efficient computation and communication overhead have been achieved by this scheme and it adequately deals with the storage in a server for few customers. [18]

The worldview of data service outsourcing is economically empowered by CLOUD computing. However, for securing data privacy, encoding of the sensitive CLOUD data must be done before outsourcing the same to the business public CLOUD and because of it, an effective data utilization service is made an extremely difficult task. [12,23,25] Users can securely search over encoded data based on keywords via conventional searchable encoded strategies, but only Boolean search are supported by them which are not adequate to meet the efficient data utilization which is characteristically in demand by enormous users and tremendous measure of files where data are stored to the CLOUD servers. [26] The problem of secured ranked keyword search algorithm over encoded CLOUD data has been solved. System usability is empowered by the ranked search algorithm by empowering search output relevance ranking as opposed to send the undifferentiated outcomes and further retrieval of the file exactness is guaranteed. [28] For guaranteeing privacy, authors implement an algorithm which is Strong Designated Verifier Signature (SDVS). [19] In this algorithm, the signers signature can be verified only by the designated verifier. [20,28,30] Another computationally indistinguishable SDVC can be produced by the designated verifier which is alluded to as non-transferability, so in the meantime the signature can't be exchanged by the designated verifier to any outsider (third party). [31] A proxy signature scheme is a unique kind of digital signature scheme among all schemes used for generating signatures where an authorized proxy signer is empowered to create a legitimate proxy signature for the benefit of the original one. As well as anyone can publicly verify the resulted proxy signature. So, a new efficient algorithm is proposed as Strong Designated Verifier Proxy Signature (SDVPS) where just a designated verifier can be persuaded of the proxy signer's identity. In an organization operation and in the electronic trade, the proposed algorithm has significant advantages. [28] Contrasted and related to the schemes, it has shorter length of the signature as well as computational expenses are lessened.

The most up to date term for the vision as long-dreamed of computing as a utility is CLOUD computing? The advantageous, on-request network is provided by the CLOUD to a centralized pool of configurable computing resources that can be deployed rapidly with higher efficiency and minimal overhead of management [4]. With its exceptional focal points, a crucial outlook change is empowered by CLOUD computing as to deploy and deliver computing services. [3,4] So, the outsourcing of the possible computing is made by it to such an extent that both individuals and an enterprise can abstain from conferring substantial costs of capital when buying and managing hardware and software, as well as additionally management of the overhead contained by operations in it.

First sub-linear searchable symmetric encryption (SSE) protocol is designed and implemented by the author where general Boolean queries and conjunctive search are

supported by it on encoded data which are symmetric. As well as it scales to vast data sets and self-assertively data those are structured which includes text search which is free. [5,10] Work in this area has concentrated essentially on searching for a single keyword. Work linear was required by earlier SSE developments where the average number of documents is stored in the database. [10,14] Great security was given by it just for attribute data which are structured, these solutions are rendered too slow and unbendable for huge databases which are practically used for the instance of conjunctive search. Conversely, authors concept gives a sensible and practical trade off amongst security and performance by proficiently supporting vast databases at the cost of all around leakage as well as moderate to the server where data are outsourced [5,7].

Problem Statement

Three entities are contained by the framework in fig. 1 as

Data Owners – which owns large set of data

Cloud Server – where the data are stored by the data owners.

Data User – who uses the data which are stored by data owners in the cloud server. Data owners grant the permission to access the data.

The responsibility of a data owner is to gather the documents, then to build an index of that document followed by outsourcing it to the server with the encryption. Aside from it, the data users have to get the approval from the data owners to access the data which are placed at the cloud server by the data owners. A numerous storage space is provided by the cloud server along with the cipher text search which needs computation assets. The encrypted index is searched by the cloud server after getting a legitimate request from the data owner and top - k documents are sent by it to the data users that are destined to coordinate data user's request. Data user chooses the number k to retrieve the documents. The framework which we have designed goes for protecting information from the spilling the data to the cloud server along with the enhancing the proficiency of cipher text search on the cloud server. The data users and the data owners are trusted in this model and the cloud server is semi trusted which means the predicted order is strictly followed by the cloud server as well as it tries to get more relevant information about the index and the data.

Threat Models - The two threat model conducts adversary's capacity. Those are as follows:

Known Cipher Text Model – document collection along with the indexed data and the keywords for querying in encrypted format can get by the cloud servers by this model.

Known Background Model – In this model more information is known by the cloud servers than the previous model i.e. Known Cipher Text Model. Information of the datasets which has the statistical background such as record recurrence as well as term recurrence information of the particular keyword can be utilized by the cloud server to dispatch a statistical assault to induce or distinguish particular keyword after querying it to the cloud server by the data users.

Goals - We have designed the system for achieving the following goals. They are as follows:

Efficient Searching – the time complexity of pursuit time should be logarithmic against the size of collection of the

data with a specific end goal with the explosive development of size of documents in enormous data situation.

Accuracy of Retrieval – Accuracy in retrieving the data from the cloud servers is identified with the two elements as the relevance between the reports which are in result set and the query which is fired to get the particular document as well as the relevance of records in the outcome set.

Search Result Integrity – The Search Result Integrity includes three perspectives as follows:

Correctness - Data owners upload the document in the Cloud server who returns the documents which remains unmodified.

Completeness – Qualified records cannot be omitted from the search results.

Freshness – The latest version of the record sets is returned from the dataset.

Protection Necessities - a progression of security necessities is set by us which now a day's researchers concentrate on.

Protection of Data - The confidentiality and the protection of records are presented by protection of data. The foe can't get the plaintext of records put away on the cloud server if information security is ensured. The ordinary approach to accomplish the data security is Symmetric Cryptography.

Privacy of an Index - Index protection implies the capacity to baffle the foe endeavour to take the data put away in the index. Such data incorporates keywords and the TF (Term Frequency) of keywords in archives, the point of records, et cetera.

Privacy of Keywords - It is imperative to ensure clients keywords which are used for querying. The algorithm which generates the query securely ought to yield trapdoor which does not leak any data regarding the keywords which are queried.

Unlink Ability for Trapdoors – Unlink ability of Trapdoors implies that each and every trapdoor produced from the query is distinctive, notwithstanding for a similar inquiry. It can be acknowledged by incorporating an irregular capacity in the trapdoor era handle. On the off chance that the enemy can derive the specific arrangement of trapdoors which all compares to a similar keyword, he can compute the recurrence of this keyword in pursuit ask for in a specific period. Consolidated with the document recurrence of keyword in known foundation model, he/she can utilize measurable assault to recognize the plain watchword behind these trapdoors.

Privacy of Rank - Rank request of indexed results ought to be very much secured. In the event that the rank request stays unaltered, the foe can look at the rank request of various indexed result, facilitate recognize the keyword search.

Accommodation of sharing the data by means of CLOUD storage, clients are likewise progressively worried about unintentional leak of data in the cloud. Such data leaks, brought about by a pernicious foe or a making trouble cloud administrator, can generally prompt to genuine breaks of individual protection or business mysteries to

address clients worries over potential data leaks in CLOUD storage.

So, primary objective of this paper is to outsource the data on the cloud servers for comfort and lessened cost in data administration. To achieve this, a secure multi-keyword ranked search algorithm has been implemented over encoded CLOUD data which supports dynamic insert and delete operations for the documents.

Proposed System

The customers are empowered for storing the data on the CLOUD servers as well as searching of the keywords is executed over cipher text area by the searchable encryption schemes. Because of various cryptography primitives, searchable encryption algorithms can be developed utilizing public key based cryptography or symmetric key based cryptography.

The principle symmetric searchable encryption was first introduced by Song et al. where each and every word of any document is encrypted by it. This strategy has a high searching expense as it scans entire data collection word by word. A protected indexed structured has been defined by Goh along with the formulation of the security model for constructing an index which is also known as semantic security against adaptive chosen keyword attack. Recently, an efficient data structure is designed and implemented by Cash et al.

A novel architecture is introduced by Cao et al. for solving the issue of searching multi keywords along with rank over encoded data which are stored on CLOUD Servers. However, the time for searching of this strategy increases exponentially going with the exponentially expanding size of the collection of document. Another design is given by Sun et al. which accomplishes better efficiency of the searching.

Disadvantages of these systems include lower search efficiency as well as no keyword privacy. So, we propose here the essential secure dynamic ranking over searching for multi-keyword algorithm in the known cipher text demonstrates. And to get the higher effectiveness the enhanced dynamic ranking over searching for multi-keyword algorithm has been used.

We propose here the essential secure dynamic multi-keyword ranked search algorithm in the known cipher text demonstrate. And to get the higher effectiveness the enhanced dynamic multi-keyword ranked search algorithm has been used. The main advantages of the proposed system are efficiency and it is secured.

The system is implemented in Fig. 2 as one of the user uploads an input file to the Cloud Server. After the file is uploaded to the server, keywords are generated from the text file and weights have been given to all the keywords using TF*IDF. Then the file and all the keywords are getting encrypted and finally, the file will be uploaded to the Cloud Storage. Then the keywords are inserted into an index.

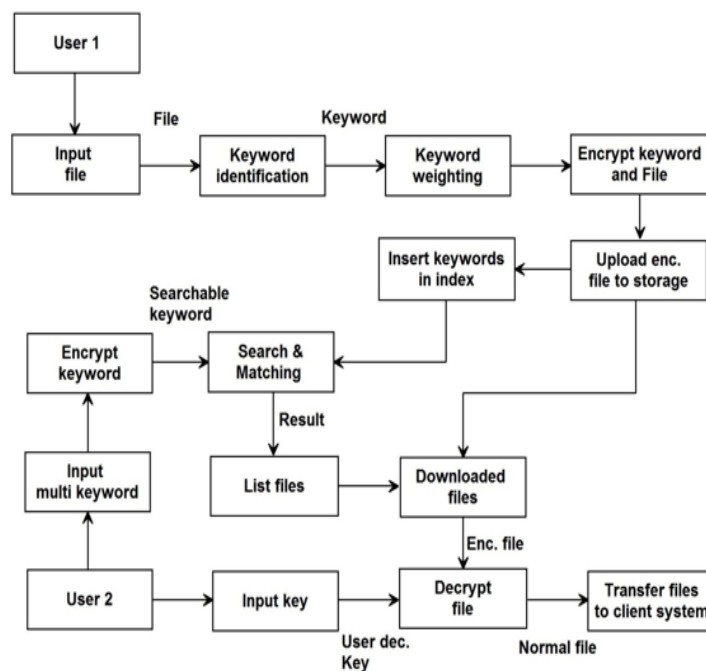


Fig. 2: System Implementation

At the same time, another user may have requested for the file. He searches for the keyword and the keyword gets encrypted and matches with all the documents which contains the keyword given for search. Then the list of files has been shown to the user which has requested for the file based on the particular keyword with the weightage of the keyword which is generated by TF*IDF. And then the file is decrypted, downloaded and transferred to the user's system who has requested for the file.

We have used the five main methodologies for constructing the proposed system as:

Key generation - In this module two keys are generated by the admin for encoding and decoding process.

Access Control - In this module access control is given by the admin for the files which he is going to upload to the CLOUD. When admin uploads the file on the CLOUD, he encodes it with the help of master secret key for the security purpose.

Keyword Indexing - In this module, the unnecessary words are removed from the file and keywords are found. Then the content weightage of the keyword has been calculated followed by converting the keywords into the hash code with the help of Message Digest 5 (MD5) scheme. And at last the hash code is placed into an indexed array.

Send Public Key - Once the user requests to the admin for the key for decryption of the particular file, the admin has to send the corresponding public key to the user's registered mail id.

Search with Keyword - Search keyword has been given as an input to the user. The keyword has been converted into a hash code. The generated hash code has been sent to the server. And on the basis of the received hash codes, server checks the index of the keywords and lists all the matching file names to the users. Then it views the shortlisted files from the server. User can download the files and the file is decrypted with the owner's public key.

In this way uploading and downloading of the file is done.

Conclusion

In this paper, the cipher text search was explored for

storing the data into the CLOUD. The issue of maintaining the semantic relation between various non-encoded documents over the related documents which are encoded is investigated and the outline technique is given to upgrade the execution search semantically. Privacy preserving is the practical issue for the data sharing system in light of public CLOUD storage where a data owner is required to circulate a substantial number of keys to the users so that they are empowered to get their documents. We proposed a concept of secure and dynamic keyword searching scheme which searches multiple keywords with rank over the data which are stored on the CLOUD in an encrypted form. Proposed architecture not only solves an issue of searching multiple keywords along with the rank associated with it but also the search effectiveness has been improved along with the rank protection as well as the relevance between the extracting documents.

References

1. S. Grzonkowski, P. M. Corcoran, and T. Coughlin, "Security analysis of authentication protocols for next-generation mobile and CE cloud services," in Proc. IEEE Int. Conf. and CE cloud services," in Proc. IEEE Int. Conf. Consumer Electron., 2011, Berlin, Germany, 2011, pp.83-87
2. D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Priv., BERKELEY, CA, 2000, pp. 44-55
3. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Proc. EUROCRYPT, Interlaken, SWITZERLAND, 2004, pp. 506-522
4. Y. C. Chang and M. Mitzenmacher, "Privacy preserving key-word searches on remote encrypted data," in Proc. 3rd Int. Conf. Applied Cryptography Netw. Security, New York, NY, 2005, pp. 442-455.
5. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved

- definitions and efficient constructions,” in Proc. 13th ACM Conf. Comput. Commun. Security, Alexandria, Virginia, 2006, pp. 79-88
6. M. Bellare, A. Boldyreva, and A. O’Neill, “Deterministic and efficiently searchable encryption,” in Proc. 27th Annu. Int. Cryptol. Conf. Adv. Cryptol., Santa Barbara, CA, 2007, pp.535-552.
 7. D. Boneh and B. Waters, “Conjunctive, subset, and range queries on encrypted data,” in Proc. 4th Conf. Theory Cryptography, Amsterdam, NETHERLANDS, 2007, pp 535-554
 8. E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol. 2003, pp. 216. 2013.
 9. C. Wang, N. Cao, K. Ren, and W. J. 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.
 10. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, “Confidentiality-preserving rank-ordered search,” in Proc. ACM ACM Workshop Storage Security Survivability, Alexandria, VA, 2007, pp. 7–12.
 11. 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., Saint Petersburg, Russia, 2009, pp. 439-449.
 12. C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, “Secure ranked key-word search over encrypted cloud data,” in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst., Genova, ITALY, 2010, pp. 253–262.
 13. P. Golle, J. Staddon, and B. Waters, “Secure conjunctive keyword search over encrypted data,” in Proc. Proc. 2nd Int. Conf. Appl. Cryptography Netw. Security, Yellow Mt, China, 2004, pp. 31–45.
 14. L. Ballard, S. Kamara, and F. Monrose, “Achieving efficient con-junctive keyword searches over encrypted data,” in Proc. 7th Int. Conf. Inform. Commun. Security, Beijing, China, 2005, pp. 414–426.
 15. R. Brinkman, “Searching in encrypted data” in University of Twente, PhD thesis, 2007.
 16. Y. H. Hwang and P. J. Lee, “Public key encryption with conjunc-tive keyword search and its extension to a multi-user system,” in Proc. 1st Int. Conf. Pairing-Based Cryptography, Tokyo, JAPAN, 2007, pp. 2–22.
 17. H. Pang, J. Shen, and R. Krishnan, “Privacy-preserving similarity-based text retrieval,” ACM Trans. Internet Technol., vol. 10, no. 1, pp. 39, Feb. 2010.
 18. N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” in Proc. IEEE INFOCOM, Shanghai, China, 2011, pp. 829–837.
 19. W. Sun, B. Wang, N. Cao, M. Li, W. Lou, Y. T. Hou, and H. Li, “Privacy-preserving multi-keyword text search in the cloud sup-ported similarity-based ranking,” in Proc. 8th ACM SIGSAC Symp. Inform., Comput. Commun. Security, Hangzhou, China, 2013, pp. 71-82.
 20. F. Li, M. Hadjieleftheriou, G. Kollios, and L. Reyzin, “Dynamic authenticated index structures for outsourced databases,” in Proc. ACM SIGMOD, Chicago, IL, 2006, pp. 121–132.
 21. H. H. Pang and K. L. Tan, “Authenticating query results in edge computing,” in Proc. 20th Int. Conf. Data Eng., Boston, MA, 2004. Pp. 560-571.
 22. C. Martel, G. Nuckolls, P. Devanbu, M. Gertz, A. Kwong, and S. G. Stubblebine, “A general model for authenticated data structures,” *Algorithmica*, vol. 39, no. 1, pp. 21–41, May 2004.
 23. C. M. Ralph, “Protocols for public key cryptosystems,” in Proc. IEEE Symp. Security Priv, Oakland, CA, 1980, pp. 122–122.
 24. R. C. Merkle, “A certified digital signature,” in Proc. Adv. cryptol., 1990, vol. 435, pp. 218–238.
 25. M. Naor and K. Nissim, “Certificate revocation and certificate update,” *IEEE J. Sel. Areas Commun.*, vol. 18, no. 4, pp. 561–570, Apr. 2000.
 26. H. Pang and K. Mouratidis, “Authenticating the query results of text search engines,” in Proc. VLDB Endow., vol. 1, no. 1, pp. 126-137, Aug. 2008.
 27. C. Chen, X. J. Zhu, P. S. Shen, and J. K. Hu, “A hierarchical cluster-ing method For big data oriented ciphertext search,” in Proc. IEEE INFOCOM, Workshop on Security and Privacy in Big Data, Toronto, Canada, 2014, pp. 559–564.
 28. S. C. Yu, C. Wang, K. Ren, and W. J. Lou, “Achieving secure, scal-able, and fine-grained data access control in cloud computing,” in Proc. IEEE INFOCOM, San Diego, CA, 2010, pp. 1–9.
 29. H. Witten, A. Moffat, and T. C. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*, 2nd ed. San Francisco, CA, USA : Morgan Kaufmann, 1999.
 30. MacQueen, “Some methods for classification and analysis of multivariate observations,” in Proc. Berkeley Symp. Math. Stat. Prob., California, 1967, p. 14.
 31. Z. X. Huang, “Extensions to the k-means algorithm for clustering large data sets with categorical values,” *Data Min. Knowl. Discov.*, vol. 2, no. 3, pp. 283–304, Sep. 1998.
 32. W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, “Secure kNN computation on encrypted databases,” in Proc. ACM SIG-MOD Int. Conf. Manage. Data, Providence, RI, 2009, pp. 139–152.
 33. R. X. Li, Z. Y. Xu, W. S. Kang, K. C. Yow, and C. Z. Xu, “Efficient Multi-keyword ranked query over encrypted data in cloud com-puting, *Futur. Gener. Comp. Syst.*, vol. 30, pp. 179–190, Jan. 2014.
 34. D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, “Highly-scalable searchable symmetric encryption with support for Boolean que-ries,” in Proc. Adv. Cryptol., Berlin, Hei-delberg, 2013, pp. 353–373.
 35. S. Kamara, C. Papamanthou, and T. Roeder, “Dynamic searchable symmetric encryption,” in Proc. Conf. Comput. Commun. Secur., 2012, pp. 965–976.
 36. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable symmetric encryption: Improved definitions and efficient construc-tions,” in Proc. 13th ACM Conf. Comput. Commun. Secur., 2006, pp. 79–88.