Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 5 – No.6, July 2016 – www.caeaccess.org

Safe Multi Keyword Ranked Search over Encrypted Cloud Data

Pradnya H. Unde
Dr.D.Y.Patil SOET,Lohagaon,
Charoli(BK), Savitribai
Phule Pune University

Arti Mohanpurkar
Dr.D.Y.Patil SOET,
Lohagaon,Charoli(BK),
Savitribai Phule
Pune University

ABSTRACT
To get more flexibility, data owner interested to outsource their complex data set on the cloud. Cloud computing provides data outsourcing and data high quality service. For data security, the data owner provides encryption on their data. Hence, for Low cost and low computational overhead the data owner migrates their data on the cloud. In subsisting system, for the first instance, the conundrum of privacy-preserving multi-keyword ranked search over encrypted data in cloud computing (MRSE) is define and solve. The basic idea of MRSE mechanism is to perform a safe inner product calculation, and then gives two appreciably improved MRSE schemes (Coordinate matching and Inner product similarity) to achieve various privacy requirements in two different models. Advance tree based index structure and numerous adaptive approaches for multi-dimensional (MD) algorithm are propose to increase the search efficiency so that the technical search efficiency is more than that of linear search.

General Terms
Multi keyword Rank search, MD Algorithm, vector Space Model.

Keywords
Cloud computing, Searchable encryption, Keyword search, Ranked.

1. INTRODUCTION
This cloud is relish as a join and Cloud computing is material which performs the computational reaction on cloud. Cloud computing technology provides benefits to easy make that are scalability, security, power, conclude as use. Benefits of outweigh are successful for providing additional facilities to client. Cloud computing provides many enrollment models i.e. proclamation as a services, the common people as a business, software as a job etc. Cloud migration is the practice of slightly or from one end to the other deploying an organization's digital capital, services, IT staple or applications to the cloud. Each world person of the house has their arrest sets of documents, to encourage those reports on their personal digital assistant or locally is esoteric process. In at variance words strengthen and collected the documents locally are valuable for computerized information purpose and it arises computational overhead. Using cloud, estate expenditure can be reduce not having to reasonable and strengthen costly hardware.

A eclipse job provider cut back deploy the front page new to their valuable performance systems, with no require to encourage and upgrade invaluable software and systems; or not exactly the employees cut back be secondhand to do some abundant work for the organization.

Hence report owner prompt to outsource their sets of documents on cloud to earn more flexibility. But before migration practice, report owner concerned to solve the report privacy issue, here after to encourage the warranty and privacy she used encryption methods. To recover the user searching practice, it is determining for one ranking system to corroborate several keywords seek, as solitary keyword search often yields easily too uncouth results.

As a common pursue indicated by today web accompany engines(e.g. Google search), announcement users might be processed to allow a set of keywords rather of only lone as the

Exasperate of their track interest to preserve the mainly data. And individually keyword in the track request is qualified to help narrow down the search result. One of the primary and widespread consist of data utilization is seek operation i.e. to all of a sudden reform information of concentration from massive amount of data. The information recovery society has the current techniques that are easily accessible to attain productive search functionalities, a well known as effect ranking and multi-keyword queries on plain text. For case, cosine solve in vector space model is a high-tech parallel measure extensively used in plain text information repossession, which involve the TF-IDF where TF stands for Term Frequency and IDF stands for Inverse Document Frequency, effort to confirm the similarity between a document and particular query, and study way precise ranked search result. However, implementing a proper description of a well known techniques during outsource encrypted keep the cloud is not approach cut is at risk privacy branch.

2. LITURATURE SURVEY
Today, Cloud computing offers a drastically contradictory and affordable act to IT resource delivery: allow the use of the announcement and processing a way with you prefer from a cloud (pool) of interconnected, given away computing systems that are maintained by Cloud function providers.

Toward Secure Multikeyword Top-k Retrival Over Encrypted Cloud Data, Jiadi Yu et al [3], Top-k Retrieval from a Confidential Index, S. Zerr, D. Olmedilla et al [10] and Fully Homomorphic Encryption over the Integers, M. van Dijk et al [11] discussed about top-k retrival methods by per confidential catalogue from one end to the other encrypted data. Homomorphic encryption techniques are secondhand for allowing persistent types of computations nearing carried inaccurate on the related cipher text.

A View of Cloud Computing, M. Armbrust et al [6] discussed approximately cloud computing benefits one as condition(agility), elasticity, availability, and cost-efficiency are amply known, discipline to asking price saving over larger economies of rocket and spongy resource appropriation schemes provided by antithetical cloud services.

Verifiable Privacy Preserving Multi-keyword Text Search in cloud supporting similarity- Based Ranking, Wenhai Sun et al [2], Secure Ranked Keyword Search over Encrypted Cloud Data, C. Wang et al [9] and Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions, R. Curtmola et al [8] and [14] discussed roughly vector space model working and humdrum based ranking multi keyword point and keyword seek scheme. Vector space epitome is respected campaign which provides tf-idf rule at the hand of which we earn accurate ranking result.

Privacy Preserving Multi-keyword Ranked Search over Encrypted Cloud Data, Ning Cao et al [1], [12] and [13] discussed close but no cigar a basic summary for the MRSE based on attain inner output computation, and before give two significantly converted MRSE schemes to get ahead distinctive resolute fig leaf requirements in two antithetical threat models. To improve attend experience of the announcement seek service, further admit of comparison with these two schemes to act as a witness more seek semantics. Thorough hit or miss investigating privacy and quickness guarantees of actual schemes is if and only if and establishes a exist of stringent privacy requirements for one a beg borrow or steal cloud story utilization system. Among various multi-keyword semantics, propose the pragmatic dreariness contrast of coordinate matching, i.e., as large amount matches as usable, to startle the relevance of data documents to the bring up the rear query. Further evaluate inner yield similarity to quantitatively manage such dreariness measure.

Cloud Migration Research: A Systematic Review, Pooyan Jamshidi et al[4] and [5] discussed roughly the power of cloud migration and the relative age of consent of this work, a everything but consolidation of existing at this moment evidence on substance to-cloud migration is timely.

Achieving Effective Cloud Search Services:Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query , Zhangjie Fu, Member et al [7] proposes an effective concern to gave a snappy comeback the check of synonym-based multi keyword ranked track from one end to

the other encrypted cloud data. The potent contributions are summarized in two aspects: synonym-based attend and similarity ranked search.

Ning Cao et al. [16] resolve retrieve ranked keyword search which exploits keyword advantage to grade results as an up to the individual of returning exact results. Wang et al. [17] used hash chain to comprise one keyword search show verification

theme. Specifically, they penetrate the applied language live clear, i.e., relevancy did a bang up job, from info retrieval to the way one sees it a win searchable index, and materialize a oneto- many order-preserving mapping technique to properly back up those sensitive did a bang up job information.

3. EXISTING SYSTEM

In cloud computing, data owner can shared their outsourced data with a number of authorized users, who might want to only retrieve the data files they are fascinated . To increase utilization or capability of access that set of data, data owner provides index keywords. Search operation is through keyword base retrieval. Keyword search is data retrieval service which applied on plain text scenarios, in which user retrieve relevant files from the sets of file based on keywords. But this scenario is becomes difficult task when it consider in the case of cipher text, because we can do only limited operations on encrypted data. To ameliorate feasibility and preserve on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most pertinent files that match users interest in lieu of all the files, which denotes that the files should be ranked in the order of pertinence by users interest and only the files with the highest pertinence are sent back to users. For the first time, existing system describe and solve the problems of multi-keyword rank search over encrypted cloud data (MRSE) while preserving system perceptive privacy in cloud computing.

In challenging, to construct the search index based on the vector space replica and adopt the cosine similarity gauge that incorporates the (TF) (IDF) weight for higher search correctness.

To improve search competence, the long vector index is split into multiple layers and suggests a tree-based index organization, where each value in anode is a sub-vector from the long index vector. Then the search algorithm is related to personalize from the MD algorithm, so as to realize more efficient search functionality.

4. PROPOSED SYSTEM

There are three main actors present in these activities: cloud server, data owner, and data user. Data owner have her own sets of documents, to maintain these documents locally is become difficult task. Maintain and stored the documents locally are expensive for storage and it arises computational overhead. Hence data owner motivate to outsource their sets of documents on cloud to get more flexibility.

But before migration process, the data privacy issue is arises in front of owner, hence to maintain the security and privacy she used encryption methods and outsource the data in encrypted form and expects the cloud server to provide keyword retrieval service to data owner himself or other authorized users. Information leakage would affect the data privacy which is unacceptable to data owner. The data user is sanctioned to process multi keyword retrieval over the outsourced data. The data user encrypts the query and sends it to the cloud server that returns the pertinent files to the data user. Afterward, the data user can decrypt and make use of the files.

4.1 Vector space Model

It is used for accurate ranking. TF-IDF rule is used to find the accurate ranking and similarity measures. Where TF denotes occurrence count of term within a document and IDF is obtained by dividing the total number of documents in collection by number of document containing the term. It gives the top-k retrieval result. IDF =total number of documents in collection/ number of documents containing the term.

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 5 – No.6, July 2016 – www.caeaccess.org

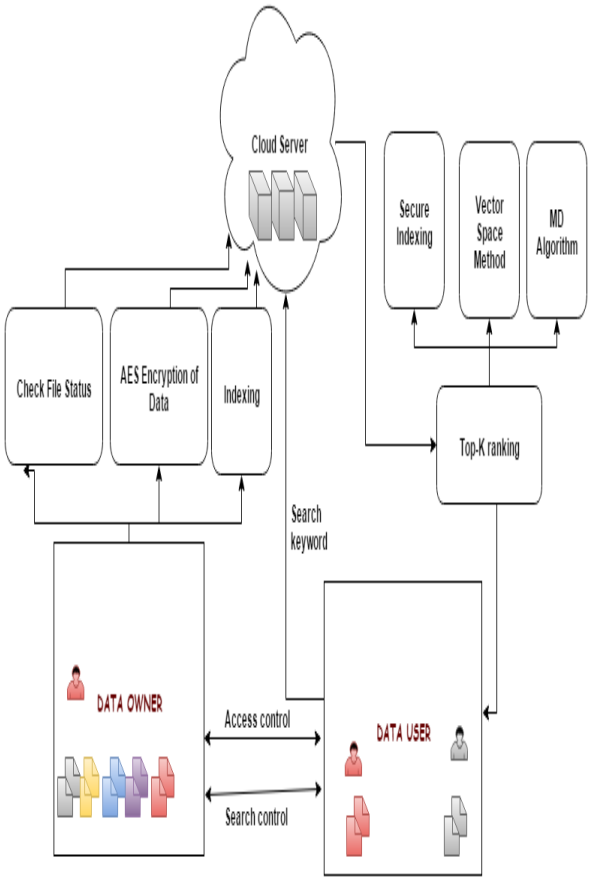

Figure 1. Architecture of proposed system

4.2 Enhance Secure Index Scheme

To achieve accurate multi-keyword ranked search, we adopt the cosine measure to evaluate similarity scores. In particular, we divide the original long document index vector into multiple sub index vectors such that each sub index represent subset of keyword and becomes a part of ith level of index tree as shown in proposed system. The query vector is divided in same way as document index vector. The final similarity score for document 'd' can be obtain by summing up the score of each level. Based on these similarity score, the cloud server determine the relevance document d to query Q and send top most relevant document to user. By using level wise secure inner product scheme, the document index vector and query index vector are both well protected.

4.3 MD Algorithm

MD algorithm is used to find k-best match in database that is structure as MDB-tree. MDB tree represents by attribute domain and each attribute in that domain has attribute value.

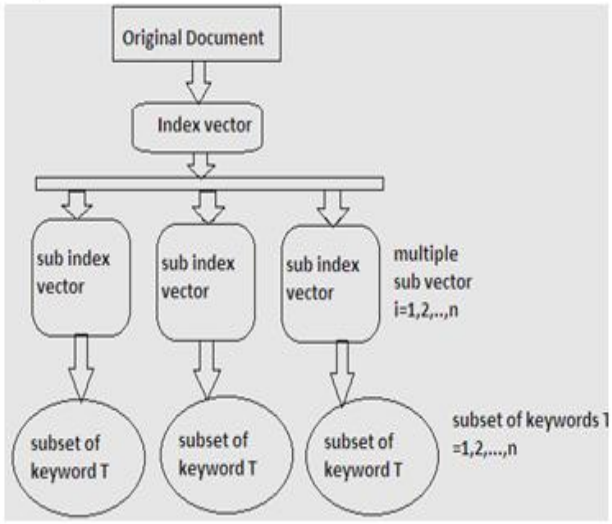

Figure 2. Mechanism of document index formation

4.4 Check File Status

Proposed system announces a third party auditor to audit user file request for checking integrity of corresponding file. Audit result from third party would be helpful for cloud service provider to enhance cloud based service platform.

5. PROPOSE SYSTEM

5.1 Setup

$P = (P_1, P_2, .., P_n)$: the plaintext document collection of n number of documents.

$E = (E_1, E_2, .., E_n)$: the encrypted document collection stored in the cloud server.

$S = (S_1, S_2, .., S_n)$: the dictionary

$I = (I_1, I_2, .., I_n)$: the searchable index associated with S.

$\widetilde{S}$ = subset of S

$T_{\widetilde{S}}$ = the trapdoor for the search request $\widetilde{S}$.

$\{M_1, M_2\}$ = {(F+1) (F+1)}

F = is the number of fields for each record r.

MRSC consist mainly four phases:

5.2 Setup Phase

Firstly, the data owner arbitrarily generates a(n+U+1)-bit vector as S and two invertible matrices $\{M_1, M_2\}$

Where n=Number of nearest neighbors,

U is the number of replica keywords implanted into every data vector.

5.3 Build Index phase

After generation of data vector, afterward, each plaintext sub index P is generated by relating dimension outspreading and splitting actions on P. Lastly, the sub index P is built for each encrypted document E.

Communications on Applied Electronics (CAE) – ISSN : 2394-4714
Foundation of Computer Science FCS, New York, USA
Volume 5 – No.6, July 2016 – www.caeaccess.org

5.4 Trapdoor phase

After this apply splitting function for Q as

$$T_{\tilde{S}} = \{M_1^{-1}\vec{Q}_i', M_2^{-1}\vec{Q}_i''\}$$

5.5 Query phase

Afterward, receiving trapdoor cloud server find similarity score. The final correspondence score calculated by cloud server is equal to

$$I_i.T_{\tilde{S}} = \{M_1^T\vec{S}_i', M_2^T\vec{S}_i''\}.\{M_1^{-1}\vec{Q}_i', M_2^{-1}\vec{Q}_i''\}$$

$$r(S_i.Q + \varepsilon_i) + t$$

The complexity of system in best case is O(n) because search time is depend on n number of keywords. And the complexity of system in worst case is O(n^2) because search time is depends on n number of keywords.

6. RESULT

Some outcomes are resulting from this scheme:

6.1 Response Time

Figure 3 shows a graph in which time require to get search result after adding number of documents in database. If database size increases then time require to get result increases.

Results must require less time for MD search as compare to MRSE technique.

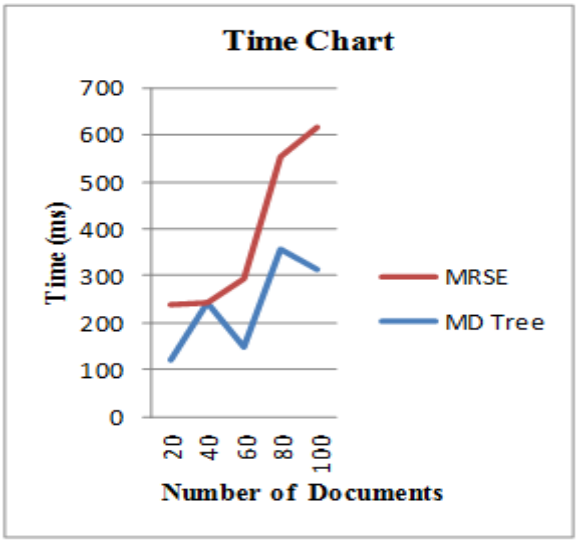

Figure 3. Response Time

6.2 Encryption time

Figure 4 shows a graph in which graph shows the expected comparative analysis for time requires to encrypt keywords using both techniques.

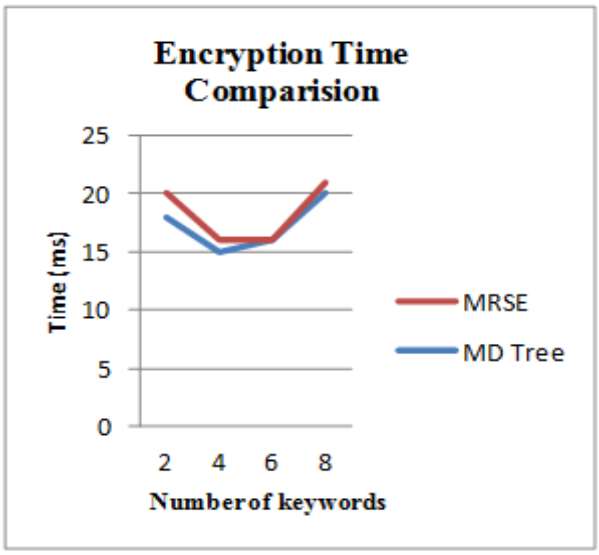

Figure 4. Encryption time comparison

6.3 Time required for trapdoor generation

Figure 5 for time require for trapdoor generation according to number of keywords present in dataset according to user search.

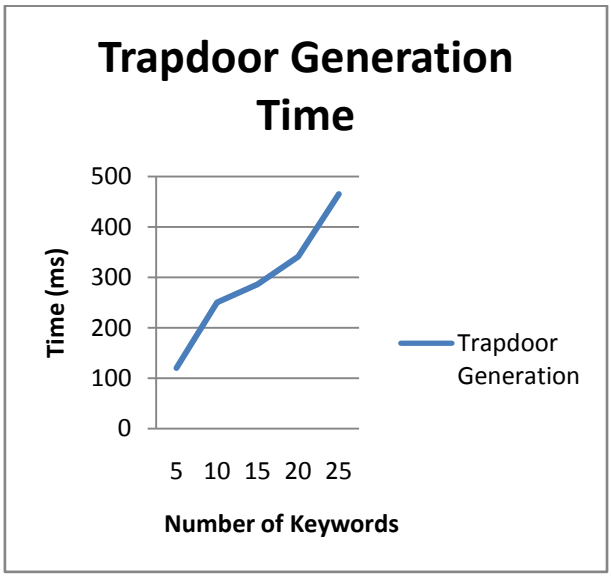

Figure 5 Time required for trapdoor generation

6.4 Results of extracted documents

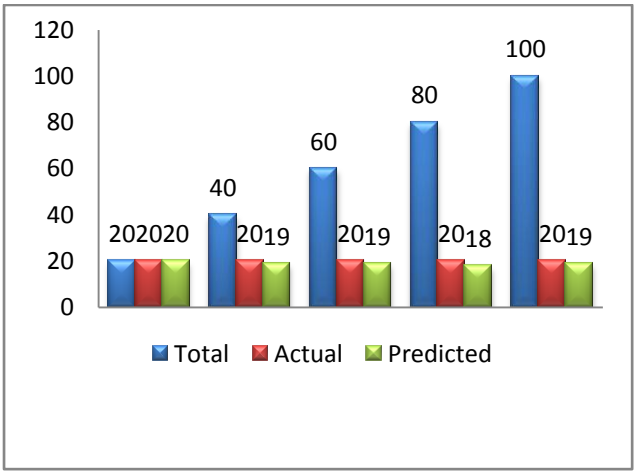

Figure 6 Results of extracted documents

Above graph shows the results of extracted (predicted) documents without giving threshold value for top-k results for keyword 'program files'. This result is analyses by increasing number of documents as 20, 40, 60, 80 and 100 for same search string 'program files'. The predicted documents by proposed system are shown in green color.

7. CONCLUSION

Cloud computing is the long dreamed vision of computing as a utility, where cloud customers can remotely store their data into the cloud so as to enjoy the on-demand high-quality applications and services from a shared pool of configurable computing resources. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plain text keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Importance of cloud migration and the relative maturity of this field, a consolidation of existing evidence on legacy to cloud migration are timely. Searchable encryption scheme is employing the fully advance encryption standards which fulfils the security requirements of multi keyword top-k retrieval over the encrypted cloud data. Proposed scheme gives guarantee of high security and practical efficiency. Future enhancement will check the integrity of rank order in the searchresult assuming the cloud server is untrusted.

8. REFERENCES

[1] Ning Cao, Cong Wang, Ming Li and Kui Ren, Privacy Preserving Multi-keyword Ranked Search over Encrypted Cloud Data, IEEE transcation on parallel and distributed system, vol. 25 no. 1, January 2014.

[2] Wenhai Sun, Bing Wang, Ning Cao, Ming Li and Wenjing Lou, Verifiable Privacy Preserving Multi-keyword Text Search in cloud supporting similarity-Based Ranking, IEEE transaction on parallel and distributed system, vol. no. 11,November 2014.

[3] Jiadi Yu, Peng Lu, Yanmin Zhu and Guangtao Xue, Toward Secure Multikeyword Top-k Retrival Over Encrypted Cloud Data, IEEE trans. on dependable and secure computing, vol. No. 4 July/august 2013.

[4] Pooyan Jamshidi, Aakash Ahmad, and Claus Pahl, Cloud Migration Research: A Systematic Review, IEEE TRANSACTIONS ON CLOUD COMPUTING, VOL. 1, NO. 2, Jul-Dec 2013

[5] V. Andrikopoulos, T. Binz, F. Leymann, and S. Strauch, How to Adapt Applications for the Cloud Environment: Challenges and Solutions in Migrating Applications to the Cloud, Computing, vol. 95, no. 6, pp. 493-535, 2013.

[6] M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, and M. Zaharia, A View of Cloud Computing, Comm. ACM, vol. 53, no. 4, pp. 50-58, 2010.

[7] Zhangjie Fu, Member, IEEE, Xingming Sun, Senior Member, IEEE, Nigel Linge, Lu Zhou, Achieving Effective Cloud Search Services :Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014

[8] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions, Proc. ACM 13th Conf. Computer and Comm. Security (CCS), 2006.

[9] C.Wang, N. Cao, J. Li, K. Ren, andW. Lou, Secure Ranked Keyword Search over Encrypted Cloud Data, Proc. IEEE 30th Intl Conf. Distributed Computing Systems (ICDCS), 2010.

[10] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, Zerber+r: Top-k Retrieval from a Confidential Index, Proc. 12th Intl Conf. Extending Database Technology: Advances in Database Technology (EDBT), 2009.

[11] Zhihua Xia, Member, IEEE, Xinhui Wang, Xingming Sun, Senior Member, IEEE, and Qian Wang, Member, IEEE, A Secure and Dynamic Multi-Keyword Ranked Search Scheme over Encrypted Cloud Data , IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS,VOL. 27, NO. 2, FEBRUARY 2016

[12] Cao, C. Wang, M. Li, K. Ren, and W. Lou, Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data, Proc IEEE INFOCOM, pp. 829-837, Apr, 2011.

[13] 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 Supporting Similarity-Based Ranking, in Proc. ACM ASIACCS, 2013, pp. 71-82.

[14] C.Wang, N. Cao, J. Li, K. Ren, andW. Lou, Secure Ranked Keyword Search over Encrypted Cloud Data, Proc. IEEE 30th Intl Conf. Distributed Computing Systems (ICDCS 10), 2010.

[15] C. Wang, Q. Wang, K. Ren, and W. Lou, Privacy-Preserving Public Auditing for Data Storage Security in Cloud Computing, Proc. IEEE INFOCOM, 2010.

[16] C. Wang, N. Cao, K. Ren, and W. Lou, Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467-1479, Aug. 2012.

[17] C. Wang, N. Cao, K. Ren, and W. 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.