International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal
ISSN No: 2456 - 6470 | www.ijtsrd.com | Volume - 2 | Issue – 3

An Efficient Fast Phrase Search with Nth-Gram For Encrypted Cloud Storage

[1]U. Sirisha, [2]R. Kalyan Kumar, [3]B. Mani Krishna

[1]Assistant Professor
[1,2,3]Dhanekula Institute of Engineering and Technology,
Ganguru, Vijayawada, Andhra Pradesh, India

ABSTRACT

Cloud computing being a new era in the research community because of its many advantages but it also consists of security and privacy concerns. The access of confidential documents and storage have been identified as one of the main issues in the area. To be precise many researchers tried to find solutions to search over encrypted documents stored on remote cloud servers. While many of them have been proposed to perform conjunctive keyword, search giving less attention on more specialized searching techniques. In this paper, we implement a phrase search technique based on Bloom filters that is significantly faster than existing solutions, with similar or better storage and communication cost. This technique uses a series of n-gram filters to support the functionality. This provides a trade-off between storage and false positive rate and is capable to defend against inclusion-relation attacks.

Keywords: *Cloud Computing, Phrase Search*

INTRODUCTION:

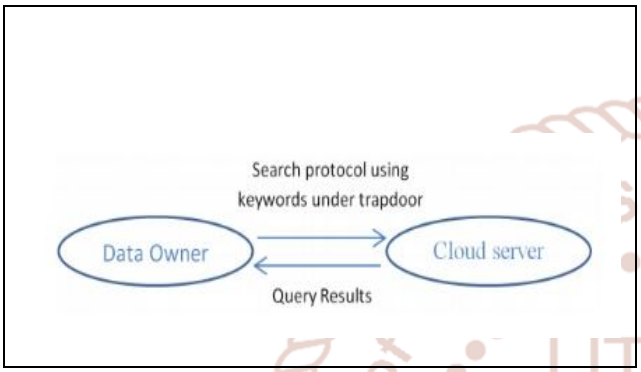
Organizations and individuals adopt cloud technologies, many have become aware of the serious concerns regarding security and privacy of accessing personal and confidential information over the Internet. The recent and continuing data breaches highlight the need for more secure cloud storage systems. While it is generally agreed that encryption is necessary, cloud providers often perform the encryption and maintain the private keys instead of the data owners. That is, the cloud can read any data it desired, providing no privacy to its users. The storage of private keys and encrypted data by the cloud provider is also problematic in case of data breach. Hence, researchers have actively been exploring solutions for secure storage on private and public clouds where private keys remain in the hands of data owners.

In this paper, we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. With modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data. Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches.

Server Usage

Keyword search framework using two parties: The data owner and an untrusted cloud server. Our algorithms can easily be adapted to the scenario of an organization wishing to setup a cloud server for its employees by implementing a proxy server in place of the data owner and having the employees/users authenticate to the proxy server. A standard keyword search protocol is shown in figure 1. During setup, the data owner generates the required encryption keys for hashing and encryption operations. Then, all documents in the database are parsed for keywords. Bloom filters tied to hashed keywords and n-grams are attached. The documents are then symmetrically encrypted and uploaded to the cloud server. To add

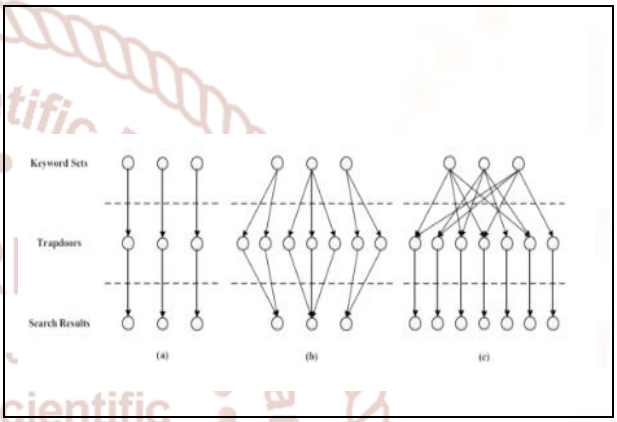
files to the database, the data owner parses the files as in setup and uploads them with Bloom filters attached to the cloud server. To remove a file from the data, the data owner simply sends the request to the cloud server, who removes the file along with the attached Bloom filters. To perform a search, the data owner computes and sends a trapdoor encryption of the queried keywords to the cloud to initiate a protocol to search for the requested keywords in the corpus. Finally, the cloud responds to the data owner with the identifiers to the requested documents.



Work Done

Boneh on an encrypted keyword search scheme based on public key encryption was among the most cited in the area. The author considered a scenario where a user wishes to have an email server verify messages associated with certain keywords without revealing the content of the emails. As sample application, the scheme would allow an urgent encrypted email to be flagged to the attention of a user while others sent to appropriate folders. The proposed solution uses identity-based encryption and a variant using bilinear mapping. Another interesting application was proposed regarding searching through encrypted audit logs, where only relevant logs are retrieved. The scenario involves an auditor which acts as a key escrow authorizing investigators to search audit records. The scheme uses an extension of Boneh's scheme using identity-based encryption. Song also considered the scenario introduced by Boneh. And proposed a probabilistic search solution based on stream cipher. Many recent works have focused on conjunctive keyword search. Ding extended Boneh scheme using bilinear mapping to perform multiple keyword search and described a solution that did not include expensive pairing operations in the encryption and trapdoor generation phase. Kerschbaum considered the search of unstructured text, where positions of keywords are unknown. The use of encrypted index for keyword search was examined in and a scheme secure against chosen keyword attack was proposed. The ranking of search results was looked at by Wang. The authors described a solution based on the commonly used TFIDF (Term Frequency X Inverse Document Frequency) rule and the use of order preserving symmetric encryption. Liuet considered the search for potentially erroneous keywords termed fuzzy keyword search. The index-based solution makes use of fuzzy dictionaries containing various misspelling of keywords including wildcards.

Relationship between keyword set, trapdoor and search result



Queries on Long Phrases

Long phrase queries are often used to locate known items rather than to locate resources for a general topic. In many cases, the goal is to identify a single document. Longer phrases also have a very low probability of occurrence and yield fewer matches. Therefore, even with a precision rate of50%, we would rarely see more than a single false positive for a search query of longer phrases. In our experiment, we never encountered more than a single false positive inqueries with phrases containing more than 4 keywords. The small number of false positives can also be easily identified and removed client-side. As a result, the effect of low precision rate in longer phrases should not have a noticeable detrimental effect in practice.

CONCLUSION

In this paper, we presented a phrase search scheme based on Bloom filter that is significantly faster than existing approaches, requiring only a single round of communication and Bloom filter verifications. The solution addresses the high computational cost by reformulating phrase search as n-gram verification

rather than a location search or a sequential chain verification. Unlike others our schemes consider only the existence of a phrase, omitting any information of its location. Being different our schemes do not require sequential verification, is parallelizable and has a practical storage requirement. Our approach is also the first to effectively allow phrase search to run independently without first performing a conjunctive keyword search to identify candidate documents. The technique of constructing a Bloom filter index enables fast verification of Bloom filters in the same manner as indexing.

REFERENCES

1. K. Cai, C. Hong, M. Zhang, D. Feng, and Z.Lv, "A secureconjunctive keywords search over encrypted cloud data againstinclusion-relation attack," in IEEE International Conference on CloudComputing Technology and Science, 2013, pp. 339–346.
2. Y. Yang, H. Lu, and J. Weng, "Multi-user private keyword searchfor cloud computing," in IEEE Third International Conference onCloud Computing Technology and Science, 2011, pp. 264–271.
3. C. Wang, N. Cao, J. Li, K. Ren, and W. Lou, "Secure ranked keyword search over encrypted cloud data," in International Conference on Distributed Computing Systems, 2010, pp. 253–262.
4. M. T. Goodrich, M. Mitzenmacher, O. Ohrimenko, and R. Tamassia, "Practical oblivious storage," in Proceedings of the Second ACMConference on Data and Application Security and Privacy, 2012, pp.13–24.