International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

An Efficient Data Search in Cloud Computing Using Search Algorithm for Ranked Algorithm

A. Boobesh[1], Pradeepa Devapriya[2]

[1]UG Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha University, Chennai, Tamil Nadu, India

Abstract: ***Cloud computing economically permits the paradigm of information service outsourcing. However, to safeguard information privacy, sensitive cloud information got to be encrypted before outsourced to the industrial public cloud, that makes effective information utilization. within the projected system, the matter of effective secure stratified keyword search over encrypted cloud information is finished. Stratified keyword search greatly enhances the system usability by returning the matching files in an exceedingly stratified order. The files ar matched consistent with sure criteria and it makes one step nearer towards the readying of privacy-preserving information hosting services in Cloud Computing. The ensuing style is ready to facilitate economical server-side ranking while not losing keyword privacy.***

Keywords: stratified keyword search, confidential information, searchable secret writing, cloud computing.

1. Introduction

Cloud Computing permits cloud customers to store their information into the cloud associated provides an on-demand top quality applications and services from a shared pool of configurable computing resources. the advantages brought by this new computing model embody however aren't restricted to: relief of the burden for storage management, universal information access with freelance geographical locations, and dodging of cost on hardware, software, and personnel maintenances, etc.

On the one hand, to fulfill the effective information retrieval would like, great deal of documents demand cloud server to perform result connection ranking, rather than returning uniform result. Such stratified search system permits information users to search out the foremost relevant data quickly. stratified search may elegantly eliminate reserve network traffic by causation back solely the foremost relevant information.

The other hand, to enhance search result accuracy in addition as enhance user looking out expertise, it's additionally crucial for such ranking system to support multiple keywords search, as single keyword search typically yields way too coarse result. As a typical follow indicated by today's internet search engines (e.g., Google search), data users might tend to supply a group of keywords rather than only 1 because the indicator of their search interest to retrieve the foremost relevant information.

Some recent styles are planned to support Boolean keyword search as an effort to complement the search flexibility, they're still not adequate give users with acceptable result ranking practicality and solves the secure stratified search over encrypted information with support of solely single keyword question. the remainder of the paper is organized as follows. within the next section, we have a tendency to gift a number of the connected add this direction. Section III describes the system design for multi-keyword search. In section IV, the small print concerning the system methodology is mentioned. Section V outlines the long run work and section VI concludes the paper.

2. Connected Work

The evolution of cloud computing is one in all the most important advances within the technologies that represent cloud computing: Platform-as-a-service (PaaS), Software-as-a-Service (SaaS) and Infrastructure-as-a-Service (IaaS).Public key coding [8] deals with the privacy of information knowledge. There area unit 2 completely different scenarios: public databases and personal databases. non-public knowledgebases: A user needs to transfer its non-public data to a far off information and needs to stay the info non-public from the remote information administrator. an extra privacy demand is to cover any data from the information administrator concerning the access pattern, i.e. if some item was retrieved over once, some item wasn't retrieved.

Public Databases: The information knowledge is public (such as stock quotes) however the user is unaware of it and needs to retrieve some data-item or hunt for some data-item, while not revealing to the information administrator that item it's. In Confidentiality-Preserving Rank-Ordered Search, once a certified user remotely accesses the info to look and retrieve desired documents, the big size of the collections typically makes it impracticable to ship all encrypted knowledge to the user's facet, then perform coding and search on the user's trust worthy computers. Therefore, new techniques area unit required to encode associated organize the info collections in an exceedingly method on enable the info center to perform economical search in an encrypted domain [4]. Order-Preserving biradial coding (OPSE), could be a settled coding theme whose coding perform preserves numerical ordering of the plaintexts. OPSE is that the kind of one-part codes, that area

unit lists of plaintexts and therefore the corresponding cipher texts, each of that area unit organized in associate alphabetical or numerical order in order that one copy is needed for economical coding and coding [5]. OPSE not solely permits economical vary queries, however additionally permits assortment and question process to be done precisely and is economical for unencrypted knowledge. knowledge owner features a assortment of the info files and that they source the info on to the cloud server in associate encrypted kind for the effective utilization of the info [1]. To do so, before outsourcing, knowledge owner can 1st build a secure searchable index from {a set|a group|a assortment} of distinct keywords extracted from the file collection, and store each the index and therefore the encrypted file assortment on to the cloud server. to look the file assortment for a given keyword, a certified user generates and submits a quest request in an exceedingly secret kind.An authorized user remotely accesses the info to look and retrieve the required documents, which frequently makes it impracticable to ship all encrypted knowledge to the user's facet, then perform coding and search on the user's trustworthy computers. Therefore, new techniques area unit required to encode and organize the info collections in such how thus on enable the info centre to perform economical search in encrypted domain [6]. the necessities of equalization privacy and confidentiality efficiently and accuracy create important challenges to the planning of search schemes for variety of search situations.

3. System Design

The summary of the projected system is illustrated in Figure one. we tend to assume that the parties ar semi-honest and don't interact with one another to bypass the protection measures the info owner uploads these search index files to the server at the side of the encrypted documents.

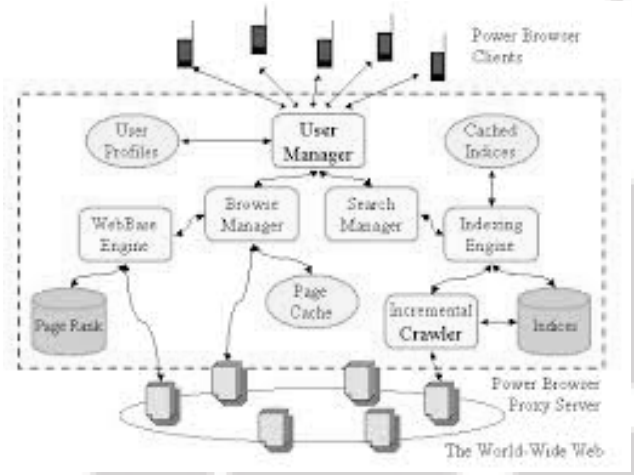


Figure 1: Multi-keyword Ranked Seacrh Architecture

The search index is formed employing a secret key based mostly trapdoor generation operate wherever the key keys square measure solely renowned by the info owner. The encoding technique will handle massive document size with efficiency.

Knowledge owner will transfer any text files on to the server, the most server can verify the index info gift in it and diverts the question to the corresponding cloud servers. the most keywords square measure extracted and therefore the unwanted words square measure filtered by stemming method. The file names square measure updated within the corresponding cloud servers. The question of the user is encrypted victimization RSA algorithm; this encoding method can stop the info larceny from the hackers. The files square measure retrieved to the user because the index knowledge of all the files square measure maintained within the index of the most cloud server.

Upon receiving knowledge, cloud server is accountable to go looking the index and come back the corresponding set of encrypted documents. to enhance document retrieval accuracy, search result ought to be graded by cloud server in line with some ranking criteria (e.g., coordinate matching, as are introduced shortly). Moreover, to cut back communication value, knowledge user might send associate nonobligatory variety so cloud server solely sends back top-k documents that square measure most relevant to the search question.

Once a user needs to perform a keyword search, he 1st connects to {the knowledge|the info|the info} owner and hunt for data while not revealing the keyword information to the info owner. The user generates the question and submits it to the server. In return, he receives data for the matched documents during a rank ordered manner.

Then the user retrieves the encrypted knowledge he chooses once analysing the data that essentially conveys a connectedness level of the every matched document, wherever the quantity of documents came is specified by the user. Finally, the user interacts with the info owner so as to decipher the documents and find the corresponding plaintext. this can be the method of multi-keyword search design.

4. System Methodology

Cloud house owners publish sure files and that they ar encrypted thus on maintain privacy. The server will hold sure keywords and therefore the method may be performed once any user accesses a specific knowledge or file. this may be done supported ranking method. The theme is shown as follows.

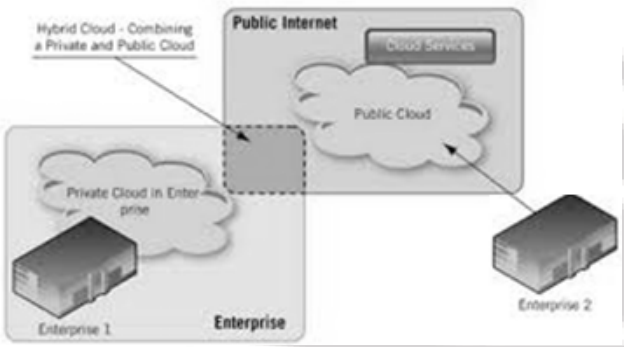
1. Setup: knowledge owner arbitrarily generates a group of files on to the cloud.
2. Build Index: The cloud may be of index server that holds data regarding the servers.
3. Trapdoor: this may be preventing varied attackers from accessing personal files exploitation encoding techniques.
4. Query: this may tend by the users to go looking for any explicit file or data

4.1 Cloud Organization and Stemming

Cloud servers square measure created with the files and therefore the index data square measure maintained within the main cloud server. Question is given to the most cloud server, so the most cloud server can verify the index data gift in it and divert the question to the corresponding cloud servers.

The words within the files square measure extracted to filter the unwanted words victimization word stemmer rule. The keywords square measure passed on to the cloud. The file names square measure updated within the corresponding cloud servers. The index server is that the main server that holds all the data of the corresponding servers.

Cloud server by choice needs to try and do thus for saving price once handling sizable amount of search requests, or there could also be package bugs, or internal/external attacks. Thus, sanctioning a hunt result authentication mechanism that may notice such sudden behaviour of cloud server is additionally of sensible interest and prices any investigation.



4.2 Encryption Technique

The question of the user is encrypted victimisation RSA algorithm; this encoding method can forestall the information felony from the hackers. Information security is ensured victimisation RSA encoding. RSA (which stands for Rivest, Shamir associated Adleman United Nations agency 1st publically delineated it) is an formula for public-key cryptography.

The encrypted scores area unit the sole extra info that somebody will utilize against the protection guarantee, i.e., keyword privacy and file confidentiality. Cloud server acts in associate honest fashion and properly follows the selected protocol specification. However, it's curious to infer and analyze information (including index) in its storage and message flows received throughout the protocol thus on learn extra info.

Due to the protection strength of the file encoding theme, the file content is clearly well protected. Thus, we have a tendency to solely got to target keyword privacy. it's the primary formula glorious to be appropriate for sign language furthermore as encoding, and was one in all the primary nice advances publicly key cryptography. RSA is wide utilized in electronic commerce protocols, and is believed to be secure given sufficiently long keys and also the use of up-to-date implementations.

4.3 Ranking and Best File Identification

Ranking the most effective file is finished by conniving the magnitude relation between the frequency and therefore the total variety of keywords. The worth is calculated and compared with the remainder of the values. The utmost valued files square measure graded so as. The files square measure retrieved to the user because the index information of all the files square measure maintained within the index of the most cloud.

The ranking is finished on the user aspect, which can usher in large computation and post process overhead. Moreover, causing back all the files consumes giant undesirable information measure. The most effective file identification is achieved victimization high k question method.

The search question is additionally represented as a binary vector wherever every bit means that whether or not corresponding keyword seems during this search request, therefore the similarity can be specifically measured by real of question vector with information vector. However, directly outsourcing information vector or question vector can violate index privacy or search privacy. The maximum graded values square measure obtained victimization term frequency calculation. The files square measure unbroken within the ascending order. The most effective files square measure given as output to the most cloud server. the most cloud server retrieves high files and given as output to the user.

4.4 Security Analysis

Cloud security architecture is only effective if the correct defensive implementations are in place. Efficient cloud security architecture should recognize the issues that will arise with security management. The security management addresses these issues with security controls. These controls are put in place to safeguard any weaknesses in the system and reduce the effect of an attack.

The cloud server should not learn the plaintext of either the data files or the searched keywords. The new scheme embeds the encrypted relevance scores in the searchable index in addition to file ID. Thus, the encrypted scores are the only additional information that the adversary can utilize against the security guarantee, i.e., keyword privacy and file confidentiality. To enable ranked search for effective utilization of outsourced cloud data under the aforementioned model, our system design should simultaneously achieve security and performance guarantees as follows.

1. Multi-keyword Ranked Search: To design search schemes which allow multi-keyword query and provide result similarity ranking for effective data retrieval, instead of returning undifferentiated results.
2. Privacy-Preserving: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements.
3. Efficiency: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

5. Future Directions

There are attainable enhancements and undergoing efforts which will seem within the future work. Firstly, the user facet of projected system are enforced on mobile devices running robot and iOS in operation systems since the potential application state of affairs envisions that users

access the info anyplace and anytime. Secondly, the projected methodology are tested on a true dataset so as to check the performance of our ranking methodology with the ranking ways utilized in plain datasets that don't involve any security or privacy-preserving techniques.

6. Conclusion

The matter of finding economical stratified keyword search is to attain the effective utilization of remotely hold on encrypted information in Cloud Computing. A basic theme shows that by following an equivalent existing searchable cryptography framework, it's terribly inefficient to attain stratified search. This befittingly weaken the protection guarantee, resort to the recently developed encrypted algorithms that permits the potency in cloud. Investigations of privacy and potency guarantees of projected schemes is given, and experiments on the real-world dataset shows however our projected schemes introduce low overhead on each computation and communication.

References

[1] Cong Wang, Ning Cao, Jin Li, Kui Ren and Wenjing Lou, sanctioning Secure and economical hierarchic Keyword Search over Outsourced Cloud information,2012.
[2] S. Kamara and K. Lauter, cryptological cloud storage, Workshop on Real-Life cryptological Protocols and Standardization 2010, Jan 2010.
[3] P. Golle, J. Staddon, and B. R. Waters, Secure Conjunctive Keyword Search over Encrypted information, 2004.
[4] Swaminathan, Y. Mao, G.-M. Su, H. Gou, A. L. Varna, S. He, M. Wu, and D. W. Oard, Confidentiality-preserving rank ordered search, 2007.
[5] Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, Order-preserving regular coding, Springer, 2009.
[6] Y.-C. Yangtze Kiang and M. Mitzenmacher, Privacy protective keyword searches on remote encrypted information, 2005.
[7] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, Top-k retrieval from a confidential index, 2009.
[8] D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, "Public key coding with keyword search, Springer, 2004.
[9] Y. H. Hwang and P. J. Lee, Public Key coding with Conjunctive Keyword Search and Its Extension to a Multi-User System, 2007.