IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified

Vol. 7, Issue 4, April 2018

Developments in Searching and Efficient Retrieval of Encrypted Data in Cloud Computing

Sunil Kumar R.M

Assistant Professor, CSE Department, Presidency University, Bangalore, India

Abstract: Cloud storage is one the most broadly used applications of cloud. As use of cloud is expanding, critical and individual data is additionally being outsourced making it imperative to keep up confidentiality and integrity of this data. A basic method for securing data is encoding it before outsourcing, yet the recovery of required records from the encrypted cloud turns into an issue which requires searching over encrypted data. Different plans have been proposed to manage this issue for searching over encrypted cloud data, and work keeps on progressing endeavoring to provide ideal user search experience resembling plaintext search. This paper reviews research in this field from single keyword to multi-keyword search, forward indexing to reverse indexing, and disjunctive to conjunctive multi-keyword search. As research in this space is developing soon with focus of influencing user search experience over encrypted data resemble plain text search experience like "Google Search".

Keywords: Multi-keyword ranked search; searchable encryption; fuzzy search; synonym search; encrypted cloud; cloud security.

I. INTRODUCTION

As data generated and stored by people and organizations is expanding exponentially over time the need to outsource nearby database systems into the cloud has become attributable to adaptability and cost benefits. To provide privacy and security to the sensitive data, we need its encryption and preceding outsourcing. This encryption represents another issue of looking over the encoded data as it obsoletes the conventional plaintext strategies for keyword search. Henceforth, enabling a search service over encrypted cloud data which is protection guaranteeing is fundamental for using cloud storage for sensitive data.

Whenever one plays out a plain text search over a database or web, the internet searcher is said to be efficient if it satisfies the criteria given below:

- Returns comes applicable to our search.
- Returns results quickly.
- Auto rectifies our spell or portrayal blunders and returns results as per possible correct keywords.
- Returns results that is relevant.

These are highlights that have been achieved for plain text and are presently a basic piece of all web search tools in any event if not database search. Fig. 1 demonstrates the four basic modules for a web index or Information Retrieval (IR) System.

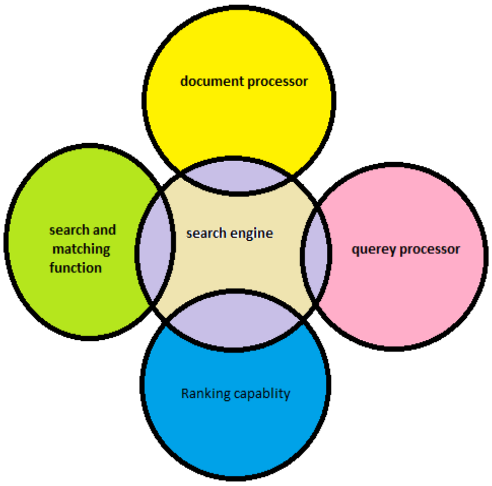

Fig. 1. Four Essential Modules Comprising

DOI 10.17148/IJARCCE.2018.7452

a Search Engine or IR System

In any case, while searching over encrypted data the situation changes and the methods produced for plain text systems are never again relevant in the encrypted data scenario. Cloud users outsourcing their data over a public cloud need to have the capacity to search and access the information as freely as plain text users.

Numerous researchers have been working towards this issue on providing users with a efficient search experience over encrypted cloud. Their exploration went for privacy protection and efficient search benefits over encrypted cloud which likewise guarantee great execution and simplicity of use. The accompanying is a review of a portion of the work in this field.

II. SINGLE KEYWORD SEARCH

Research in the field of searching over encrypted cloud data began with single keyword search.

Song et al. [1] was first to suggest searchable encryption. Their proposition was to encrypt each word of the archive independently which however brought about expanded cost as all reports of the dataset must be examined word by word, rendering the scheme wasteful. They proposed a consecutive sweep could be performed with or without a index. The pros of utilizing a index is the point at which the records in the dataset are genuinely substantial and is such circumstances the file yields quicker pursuit. The cons include storage and updating overhead to which the use of index is better suited to read-only data collections.

Goh et al. [2] proposed a index architecture which guaranteed security. The protected index structure enabled a trapdoor comparing to a keyword K to be checked in O(1) time if K was a piece of the index. In this procedure it was likewise guaranteed that no extra information can be gotten from the record In any case, while searching over encrypted data the situation changes and the methods produced for plain text systems are never again relevant in the encrypted data scenario. Cloud users outsourcing their data over a public cloud need to have the capacity to search and access the information as freely as plain text users.

Numerous researchers have been working towards this issue on providing users with a efficient search experience over encrypted cloud. Their exploration went for privacy protection and efficient search benefits over encrypted cloud which likewise guarantee great execution and simplicity of use. The accompanying is a review of a portion of the work in this field. Trapdoors are secure as they are made utilizing a secret key. Semantic security against adoptive chosen key attack (ind-cka) was their security model and they likewise figured ind-cka secure index development by the name z - idx which was effective as it used Bloom filters alongside pseudo-random functions. These techniques created by them can be utilized for applications, for example, developing encrypted audit logs and other database query schemes where privacy must be guaranteed.

Wang et al. [3] proposed a secure ranked keyword search over encrypted cloud data. Ranked search method improves framework convenience, as it were, as in it restores the coordinating records in a positioned arrange in view of certain important criteria like watchword recurrence and so on., along these lines making the pursuit over cloud more easy to understand. Wang et al. proposed a plan under accessible symmetric encryption (SSE) security, and went ahead to demonstrate how it was wasteful. To improve common sense they at that point proposed arrange - protecting symmetric encryption (OPSE) which gave more grounded security ensure in contrast with SSE plans, alongside accomplishing the motivation behind ranking results.

Wang et al. [4] went ahead to propose a encrypted invert index. This was gone for accomplishing secured ranked search over encrypted cloud data by computing relevance score amongst reports and qrery, so as to rank relevent documents as indicated by the relevance score and empower client recover the most pertinent n results.

Cash et al. designed and implemented an efficient data structure. It was frequently watched that undue user time is squandered in choosing the desired information when various outcomes are returned. This wastage of time is owing lack of ranking mechanism. Hence, order-Preserving methods were introduced and utilized [5-7].

Boneh et al. [8] were the ones to pioneer searchable encryption construction in public setting. In their plan one uses general public key to store data on the cloud yet to search private key is fundamental. They presented the idea of public key encryption nearby keyword search and gave a few constructions.

Be that as it may, all the above Techniques bolster just a single keyword search.

DOI 10.17148/IJARCCE.2018.7452

III. MULTI – KEYWORD SEARCH

Numerous conjunctive keyword search techniques have been proposed to advance search predicates. In any case, an imperfection is that these methods have expansive overheads of communication cost by sharing secret [9], or computational cost by bilinear map [10]. A conjunctive search method recommended in [11] was additionally improved by Handa et al. [12] utilizing clustering. Be that as it may, as the search is conjunctive subsequently comes about returned incorporate just the group which contains an all the search keywords. For a partial match no cluster is chosen because of which search results are not easy to use and pass up a great opportunity for multiple outcomes.

Pang et al. [13] proposed a vector space model based secure search scheme. In any case, attributable to neither conceivable security analysis of information on recurrence nor the search scheme performance, all things considered, situation, the plan's security and efficient is not clear.

Cao et al. [14] were pioneers to characterize multi-keyword ranked search problem. Their novel design for multi-keyword ranked search over encrypted cloud data handled the critical issue of privacy-preserving multi-keyword ranked search over encrypted cloud information (MRSE). They used similitude measure of "coordinate matching" as the multi-keyword semantics. This empowered them to measure relevance of documents to user input search keywords. They quantitatively assessed similarity measure using “inner product similarity”. Nonetheless, the drawbacks of this scheme were – first the static dictionary reference which must be reconstructed at whatever point new keywords were included. Also, however this was a decent approach yet exponential development in the search time with exponential increment in size of the document accumulations is a noteworthy disadvantage.

Following MRSE Cao et al. [15] proposed an enhanced scheme "Efficient Multi-Keyword Rank Query on Encrypted Cloud Data" (MKQE) [15]. MKQE's effectiveness lied in the way that it diminished the quantity of computations required on development of keyword lexicon. Another change was that it account access frequencies of keywords. However this scheme was missing as it did exclude users experience enhancing features, for example, semantic and fuzzy keywords.

Sun et al. [16] presented another architecture for searching to get enhanced efficiency yet this scheme did not handle importance between documents while creating the index which prompted restricted search results. For instance: if a client query comprised of keywords like given a query containing keywords, for example, electronic and shop at that point reports having just keywords will be returned, however considering documents containing electronic and store would be important and including such outcomes would upgrade user search experience.

IV. ENHANCING USER'S SEARCH EXPERIENCE

Once the issues concerning single and multi-keyword search were handled the subsequent stage in line was to consider user desires while searching over encrypted cloud data.

Li et al. [17] were the initial ones to give a fuzzy keyword search conspire for encrypted cloud data. The plan handled minor grammatical mistakes and configuration inconsistence first of all however did not improve user search experience by means of considering. They clarified that grammatical errors was an exceptionally normal user look marvel and by overlooking it the conventional plans restored no outcomes for each wrongly spelled outcome which decreased user search involvement and made it frustrating to use the search systems. Subsequently, Li et al. used fuzzy keywords to manage this issue while keeping up keyword privacy. Fuzzy keyword utilize alter separation to process the nearest coordinating keyword if there should be an occurrence of a user mistake. They likewise developed a propelled procedure for developing these keyword word dictionaries which reduce storage overhead close by handling tackling representation issues. They showed that the proposed plot was protection saving and also secure through thorough privacy preserving and furthermore precisely accomplished the advantages of fuzzy keyword search.

Khan et al. [18] gave a fuzzy keyword search scheme that enhanced the previous work in this area by including ranking along with multi-keyword search on the encrypted cloud thereby upgrading user search experience. They formulized and solved the issue of effective secure ranked fuzzy multi-keyword search over outsourced encrypted cloud data (RFMS). This work, being among a portion of the initial in this field, who used fuzzy search on multi keyword search alongside using positioning which enhanced pertinence of results recovered. The plan considered pertinence of results in light of the quantity of keywords coordinated. For instance if user input question contained 5 keywords out of which 2 were incorrectly spelled then the plan would first right the incorrectly spelled keywords utilizing alter separation and after that while returning outcomes the records containing each of the 5 keywords would be best of the rundown took after by those containing 4 keywords out of the 5 et cetera. The plan did not further inside rank the outcomes, which implies that it didn't consider add up to event of the 5 keywords and did not further the sort comes about where 5 keywords were coordinated by this check. Additionally, the plan didn't consider equivalent word look criteria and expanded inquiry time radically.

Chai et al. [19] proposed search plot that was undeniable and they demonstrated its proficiency alongside the culmination and accuracy of indexed lists. In any case, the disadvantages were that there were a couple of security issues which were not tended to in the work. Wang et al. [20] gave a plan that upheld fuzzy pursuit alongside confirmation in view of VSSE

(Undeniable symmetric accessible encryption), however the plan disregarded positioning of results.

Fu et al. [21] gave a plan that went ahead to consider the benefit of equivalent word question in a encrypted cloud condition which bolstered multi-keyword search. The primary commitments can be abridged into two key focuses: similitude equivalent word based pursuit and positioning of results. This plan upheld equivalent word question which manages another key user search conduct where the user may enter comparable importance words for the existent lexicon keywords rather than the coordinating keywords or fuzzy keywords as a result of either the user overlooking the correct keyword prompting equivalent word substitution or because of the user not being knowledgeable with the correct information. They utilized the vector space display alongside cosine measure (generally acknowledged strategies in the field of data recovery), to ascertain comparability between user input keywords and the dataset. In conclusion, they played out a broad and nitty gritty execution examination of the proposed plots by checking for precision and effectiveness of the hunt through trials on the dataset which constituted genuine reports. The execution examination affirmed the plan to be productive and furthermore that it bolsters equivalent word based looking successfully. In the space of searching over encoded information, this is a decent advance forward. Be that as it may, the issues of confirmation of query items and semantic hunt are not tended to in this plan and neither does it bolster fuzzy keywords.

Next Fu et al. [22] execute a savvy cloud search plot which is composed with a shopper driven view point in which user experience is of most extreme significance, and spotlight on how bolster evidence of output while guaranteeing adaptability of encrypted cloud data. They proposed a shrewd semantic hunt plot which underpins the idea of unquestionable status of query output. Additionally, the plan returns consequences of both keyword based correct matches alongside aftereffects of keyword based semantic matches. They performed thorough security and execution investigation to demonstrate that the proposed conspire successfully accomplishes the objective of keyword based semantic hunt and is secure.

Mittal et al. [23] center the level of user look experience improvement achievable by clubbing strategies of equivalent word based pursuit with fuzzy multi-keyword positioned search over encrypted cloud data, which is an issue that no existent plan has considered till date. They have proposed Privacy Preserving Synonym Based Fuzzy Multi-Keyword Ranked Search over Encrypted Cloud Data (SFMRS), a plan which upgrades user look involvement to a fundamental by giving multi-keyword positioned search utilizing both fuzzy and equivalent word keywords, in this way taking encoded look experience nearer to free content web indexes.

The plan furthermore enhanced list age time and inquiry time in contrast with existing plans by using a double tree based dynamic record. Test comes about depicted the adequacy of the plan as it decreases the pursuit time, i.e. the ideal opportunity for finding the coveted records, by 90% alongside lessening to a base the overhead to refresh the file when new documents are included the archive accumulation (index generation time) as compared to the existing efficient indexing schemes in literature for a similar dataset. Optimization of search time together with index generation time had not been achieved to this extent before. The scheme however doesn’t deal with semantic queries.

TABLE I shows the comparison in Index generation time between SFMRS and other schemes. The search times taken by SFMRS in comparison to RFMS is shown by TABLE II.

The simulations for SFMRS were finished utilizing PHP. The client side application was executed utilizing PHP, MySQL, and Apache Server over Windows7 OS on Intel i3 processor (2GB RAM). The cloud was executed on an Openshift outfit (with arrangement 1 Core processor, 1GB capacity and 512MB RAM) using PHP over Linux OS. The dataset estimate was kept like RFMS adding up to 2000 documents with 8000 one of a kind watchwords. The dataset comprises of content records (normal size 20KB) some of which are engineered information.

TABLE I: INDEX GENERATION TIME

Number of keywords in dictionary for 2000 files	Index Generation Time (s)			
	MRSE	MKQE	RFMS	SFMRS
2000	46.9	35.3	20.09	10.49 x 10^{-6}
8000	648.8	447.4	45.67	201.1 x 10^{-6}

TABLE II: TIME TAKEN TO SEARCH AND RETURN RANKED RESULTS FOR USER INPUT KEYWORDS USING SFRMS VS RFMS

Number of keywords in search query	Search Time (ms)							
	100 keywords and 100 files		1000 keywords and 1000 files		4000 keywords and 2000 Files		8000 Keywords and 2000 files	
	SFRMS	*RFMS*	*SFRMS*	*RFMS*	*SFRMS*	*RFMS*	*SFRMS*	*RFMS*
1	1.28	97.72	3.42	303.84	4.75	489.65	7.04	259.16
2	2.90	90.34	7.74	283.84	10.74	345.65	15.93	312.38
3	5.98	96.04	15.99	148.74	22.18	451.04	32.90	309.77
4	4.03	94.90	10.77	233.51	14.93	286.07	22.15	249.16
5	4.18	85.21	11.17	317.56	15.50	312.38	22.99	309.77
6	4.74	82.89	12.68	189.04	17.59	309.77	26.09	309.07
7	5.96	90.34	15.92	322.61	22.08	256.07	32.76	356.98
8	4.79	90.34	12.81	198.75	17.77	271.04	26.36	269.16
9	6.55	85.97	17.51	188.09	24.29	286.07	36.03	243.16
10	4.79	81.74	12.80	207.67	17.76	312.38	26.35	269.16

V. CONCLUSION

Work has been done toward giving upgraded and effective watchword seek understanding to clients over encoded cloud. Be that as it may, distinctive techniques have diverse downsides or confinements. Some address simply single catchphrase inquiries and out of those positioning isn't considered in a large portion of them. Papers which use conjunctive strategies or disjunctive systems to perform multi-watchword seek have restricted flexibility and proficiency as far as meeting client desires. Subsequently, writing demonstrates that in spite of accomplishments, scope for development exists and client look experience can be additionally improved. A few fields in which work is at present being done is semantic pursuit on scrambled information of an outsourced cloud. Analysts have chipped away at different features autonomously and if investigate in this space continues developing at this pace then ideally soon client seek involvement over scrambled information will look like plain content inquiry encounter, (for example, "Google Search").

REFERENCES

[1] D. X. D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in IEEE Symposium on S & P, pp. 44-55, 2000.
[2] E.-J. Goh, Secure Indexes, IACR Cryptology ePrint Archive, vol. 2003, pp. 216, 2003.
[3] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," in 30th International Conference on Distributed Computing Systems, pp. 253-262, 2010.
[4] C. Wang, N. Cao, K. Ren, and W. J. Lou, Enabling Secure and Efficient Ranked Keyword Search over Outsourced Cloud Data, IEEE Transaction on Parallel Distribributed System, vol. 23, no. 8, pp. 1467-1479, 2012.
[5] A. Swaminathan, Y. Mao, G. M. Su, H. Gou, A. Varna, S. He, M. Wu, and D. Oard, "Confidentiality-Preserving Rank-Ordered Search," in ACM StorageSS, pp. 7-12, 2007.
[6] S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, "Zerber+R: top-k retrieval from a confidential index," in EDBT, pp. 439-449, 2009.
[7] K. Li; W. Zhang; C. Yang; N. Yu, " Security Analysis on One-to-Many Order Preserving Encryption-Based Cloud Data Search", IEEE Transactions on Information Forensics and Security, vol. 10, no. 9, pp. 1918 – 1926, 2015
[8] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in EUROCRYPT, pp. 506-522, 2004.
[9] L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in ICICS, pp. 414-426, 2005.
[10] D. Boneh, and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in TCC, pp. 535-554, 2007.
[11] Cengiz Orencik and Erkay Savas, "An efficient privacy-preserving multi-keyword search over encrypted cloud data with ranking," in Springer Distributed and Parallel Databases, pp. 119–160, 2014.

DOI 10.17148/IJARCCE.2018.7452

[12] Handa, R. and Challa, R.K., “A cluster based multi-keyword search on outsourced encrypted cloud data,” International Conference on Computing for Sustainable Global Development (INDIACom), pp. 115-120, 2015.

[13] S. Som, S. Sinha, R. Kataria (2016) “STUDY ON SQL INJECTION ATTACKS: MODE, DETECTION AND PREVENTION”, International Journal of Engineering Applied Sciences and Technology, Indexed in Google Scholar, ISI etc., Impact Factor: 1.494, Vol. 1, Issue 8, ISSN No. 2455-2143, Pages 23-29, June - July 2016.

[14] H. Pang, J. Shen, and R. Krishnan, Privacy-Preserving Similarity-Based Text Retrieval, ACM Transactions on Internet Technology (TOIT), vol. 10, no. 1, pp. 39-42, 2010.

[15] N. Cao, C. Wang, M. Li, K. Ren, and W. J. Lou, ”Privacy-Preserving Multi-keyword Ranked Search over Encrypted Cloud Data,” in IEEE INFOCOM, pp. 829-837, 2011.

[16] Z. Xu, W. Kang, R. Li, K. Yow and C. Xu, “Efficient multi-keyword rank query on encrypted cloud data”, 18th IEEE International Conference on Parallel and Distributed Systems, pp: 244-251, 2012.

[17] 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 ASIACCS, pp. 71-82, 2013.

[18] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. J. Lou, “Fuzzy keyword search over encrypted data in cloud computing,” IEEE INFOCOM 2010, pp. 1-5, 2010.

[19] N. Khan, C. R Krishna, A. Khurana, “Secure Fuzzy Multi-Keyword Search over Outsourced Encrypted Cloud Data”, IEEE International Conference on Computer and Communication Technology (ICCCT), pp. 241 – 249, 2014.

[20] Q. Chai and G. Gong, “Verifiable Symmetric Searchable Encryption for Semi- Honest-but-Curious Cloud Servers,” IEEE International Conference on Communications (ICC’12), pp. 917-922, 2012.

[21] J. Wang, H. Ma, and Q. Tang, “A new efficient verifiable fuzzy keyword search scheme,” Journal of Wireless Mobile Networks, Ubiquitous Computing and Dependable Applications, vol. 3, no. 4, pp. 61-71, 2012.

[22] Z. Fu, X. Sun, N. Linge, and L. Zhou, “Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud

[23] Data Supporting Synonym Query,” IEEE Transaction Consumer Electronics, vol. 60, no.1, pp. 164-172, 2014.

[24] Z. Fu, X. Sun, N. Linge, and L. Zhou, “Smart cloud search services: verifiable keyword-based semantic search over encrypted cloud data,” IEEE Transaction Consumer Electronics, vol. 60, no.4, pp. 762 - 770, 2014.

[25] S. Mittal, C. R Krishna, “Privacy Preserving Synonym Based Fuzzy Multi-Keyword Ranked Search over Encrypted Cloud Data”, IEEE

[26] International Conference on Computing Communication and Automation, In Press, 2016.

DOI 10.17148/IJARCCE.2018.7452