International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Improve Efficiency for Data Privacy Keywords Using Top-K Search Scheme

V. Likhitha[1*], Mr. A. Murali Mohan Kumar[2]
(Master of Computer Applications)[1] (Associate Professor)[2]
Mother Theresa Institute of Computer Applications, Palamaner, S.V. University, Tirupathi, AP, India

Article Info

Publication Issue :
Volume 8, Issue 5
September-October-2022

Page Number : 261-266

Article History
Accepted: 10 Oct 2022
Published: 25 Oct 2022

ABSTRACT

Cloud computing provides people and enterprises huge computing power and climbable storage capacities to support a variety of huge information applications in domains like health care and research project, so additional and additional information owners are concerned to source their information on cloud servers for nice convenience in information management and mining. However, information sets like health records in electronic documents sometimes contain sensitive info, that brings regarding privacy issues if the documents are free or shared to partly untrusted third-parties in cloud. A sensible and wide used technique for information privacy preservation is to encode data before outsourcing to the cloud servers, that but reduces {the information|the info|the information} utility and makes several ancient data analytic operators like keyword-based top-k document retrieval obsolete. during this paper, we have a tendency to investigate the multi-keyword top-k search drawback for big encoding against privacy breaches, and conceive to establish an economical and secure answer to the current drawback. Specifically, for the privacy concern of question information, we have a tendency to construct a special tree-based index structure and style a random traversal algorithmic program, which makes even constant question to provide completely different visiting ways on the index, and may conjointly maintain the accuracy of queries unchanged under stronger privacy. For raising the question potency, we have a tendency to propose a bunch multi-keyword top-k search theme supported the thought of partition, wherever a bunch of tree-based indexes are created for all documents. Finally, we have a tendency to mix these strategies along into an economicaland secure approach to handle our projected top-k similarity search. intensive experimental results on real-life information sets demonstrate that our projected approach will considerably improve the potential of defensive the privacy breaches, the measurability and the time potency of question process over the progressive strategies.

Keywords: Cloud computing, privacy preserving, data encryption, multi-keyword top-k search, random traversal.

Copyright: © the author(s), publisher and licensee Technoscience Academy. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited

I. INTRODUCTION

Cloud processing has risen as a problematic pattern in both IT businesses and research groups as of late, its remarkable attributes like high versatility and pay-as you-go mold have empowered cloud buyers to buy the effective processing assets as administrations concurring to their real necessities, with the end goal that cloud clients have never again need to stress over the squandering on processing assets and the multifaceted nature on equipment stage administration. These days, an ever-increasing number of organizations and people from a substantial number of huge information applications have outsource their information and convey their administrations into cloud servers for simple information administration, proficient information mining and question handling assignments. Yet, when the organizations and people appreciate these preferences in distributed computing, they likewise need to take the security worry of the outsourced information into account. Since informational collections in numerous applications frequently contain touchy data like messages, electronic wellbeing records and money related exchange records, when the information proprietor outsourcing such delicate information to the cloud servers which are considered to be in part believed, the information can be effectively gotten to what's more, examined by cloud specialist organizations unlawfully. Since the examination of these informational collections may give significant bits of knowledge into various key territories in the public arena, (for example, e-examine, human services, restorative and taxpayer driven organizations), subsequently information proprietors require compelling, versatile and security safeguarding administrations before discharging their information to the cloud. Information encryption has been generally utilized for information security conservation in information sharing situations, it alludes to numerical count and algorithmic plan that change plaintext into cyphertext, which is a non-intelligible shape to unapproved parties. An assortment of information encryption models has been proposed and they are utilized to scramble the information before outsourcing to the cloud servers. Be that as it may, applying these methodologies for information encryption normally cause gigantic cost as far as information utility, which makes customary information handling strategies that are intended for plaintext information never again function admirably finished scrambled information. The catchphrase-based hunt is such one broadly utilized information administrator in numerous database and data recovery applications, and its conventional handling techniques can't be straightforwardly connected to encoded information. In this manner, how to process such questions over encoded information and at the same time ensure information security turns into a hot research theme. Luckily, numerous philosophies in view of accessible encryption have been contemplated. For instance, manage the single watchword pursuit, and works bolster the multi-catchphrase boolean hunt. Be that as it may, the single watchword seek isn't savvy enough to bolster propelled questions and the boolean pursuit is impossible since it causes high correspondence cost. Along these lines, later works like center around the multikeyword positioned seek, which is more down to earth in pay-as you-go cloud worldview. Be that as it may, the majority of these strategies can't meet the high inquiry productivity and the solid information security at the same time, particularly while applying them to huge information encryption postures incredible adaptability and productivity challenges. Spurred by this, in this paper, we center around a

unique sort of multi-catchphrase positioned seek, to be specific the multikeyword top-k seek, which has been an exceptionally well-known database administrator in numerous essential applications, and as it were requirements to restore the k archives with the most elevated significance scores. For supporting multi-watchword seek, we present the vector space show which speaks to reports and questions as vectors. With a specific end goal to help top-k look, the importance scores amongst records and questions ought to be figured, in this manner, the TF×IDF (term recurrence × converse record recurrence) demonstrate is presented as a weighting administer to register the significance scores for positioning purposes. What's more, to enhance the inquiry proficiency for better client encounters, we propose a gathering multi-catchphrase topk seek plot (GMTS), which depends on segment and bolsters top-k closeness look over scrambled information. In this conspire, the information proprietor separates the catchphrases in the word reference (assume that the lexicon contains every one of the watchwords that could be extricated from all reports) into numerous gatherings and builds up an accessible file for each gathering. On the opposite side, to better control the span of lists, we receive champion records into our plan, where the record of a watchword aggregate just stores the best ck reports of the relating watchword (the best ck archives of a watchword speak to the c ∗ k archives that have the most astounding pertinence scores to this catchphrase, where c is a positive whole number). Besides, we propose an arbitrary traversal calculation (RTRA) to fortify the information security, where the information proprietor fabricates a twofold tree as accessible record and allots an arbitrary change to every hub, so the information client can allot an arbitrary key to each question. Subsequently, the information client can change the outcomes and going by ways of questions by utilizing distinctive keys, which keeps up high precision of questions. At long last, we consolidate the GMTS and the RTRA together into a proficient and secure answer for our proposed issue. Our commitments can be abridged as takes after:

We initially propose the arbitrary traversal calculation which makes the cloud server haphazardly cross on record and returns diverse outcomes for the same inquiry, and meanwhile, it keeps up the exactness of inquiries unaltered for higher security. Based on the arbitrary traversal calculation, we introduce one both proficient and secure accessible encryption conspire, which can bolster top-k closeness seek over encoded information. In this plan, the information proprietor can control the level of question unlikability without giving up exactness.

Algorithm:

we propose an irregular traversal calculation (RTRA). In RTRA, giving two indistinguishable questions, their meeting ways in file and list items can be unique, however keep up the exactness of inquiries unaltered. The fundamental thought is as per the following: 1) extending the entire record accumulation E times, subsequently each archive in result has E alternatives; 2) allocating a switch to each archive; 3) fabricating a tree-based list for the entire archive gathering, where record identifiers are put away in leaf hubs. 4) allocating an arbitrary key to each question. In this way, information clients can control the meeting ways furthermore, indexed lists by allotting diverse keys. Our objectives contain three angles: 1) Supporting multikeyword top-k likeness look over scrambled information; 2) Inquiry with high proficiency; 3) Privacy-saving. The points of interest are recorded as beneath:

Multi-watchword top-k Search: To plan an accessible encryption conspire that empowers the cloud server to help multi-watchword top-k similitude seek over encoded information;

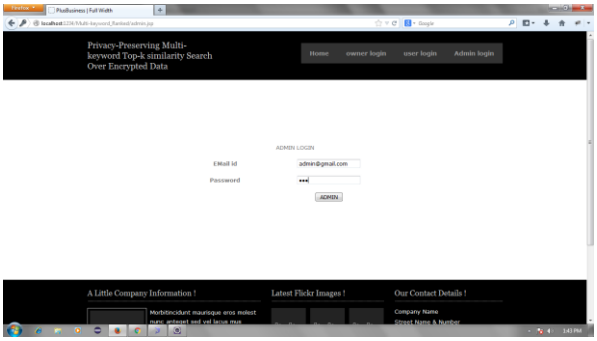
Pursuit productivity: Our plan ought to be proficient in file development, trapdoor age and pursuit preparing, furthermore, it ought to be more productive and compelling than the best-in-class techniques;

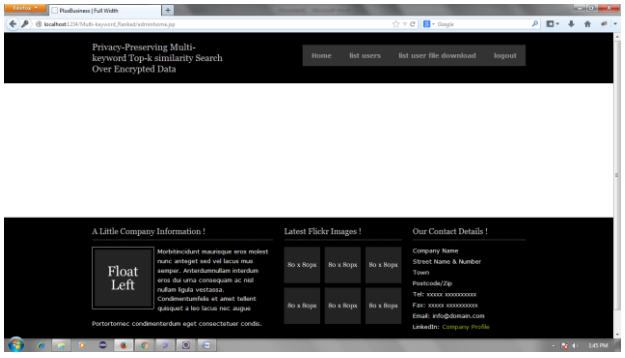
Security saving: Our plan ought to ensure the protection of lists and inquiries in the meantime. They are Index security and Query security: The plaintext data of scrambled accessible list and trapdoor ought to be secured.

Keywords Privacy: The cloud server can't recognize regardless of whether a specific catchphrase is contained in a question by breaking down records or query items.

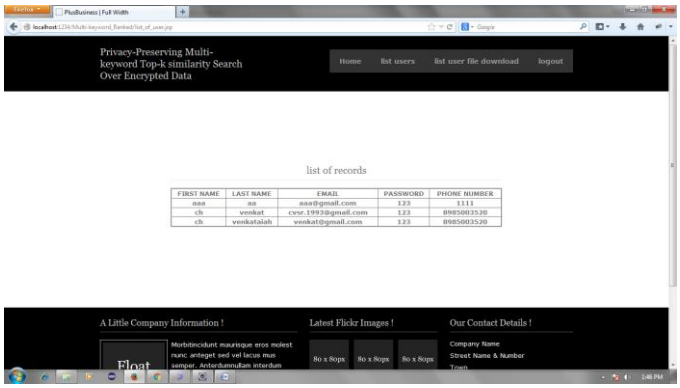
Results:

Admin login

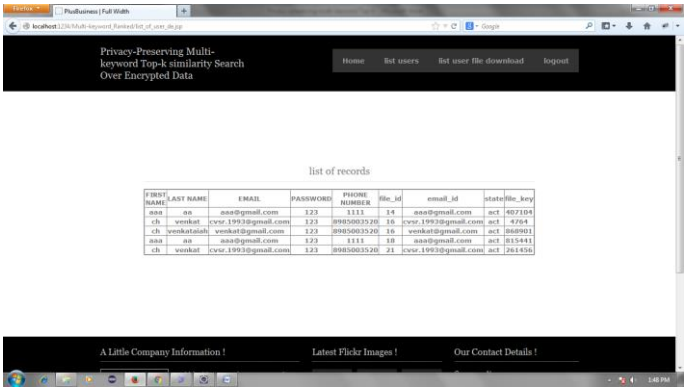




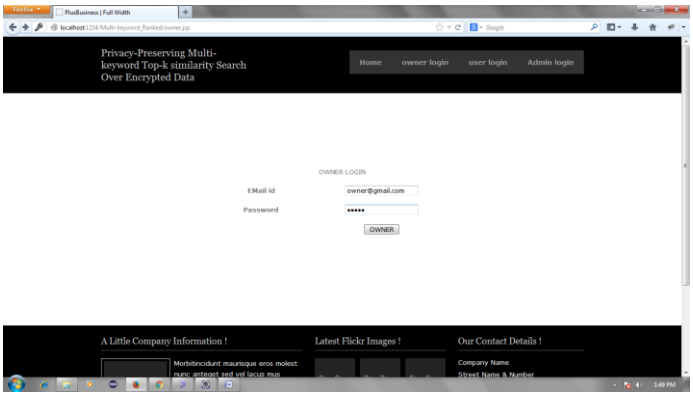
Click→list users



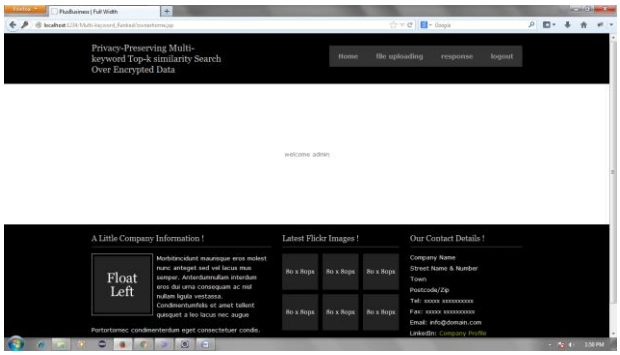
Click→list user file download



Owner login



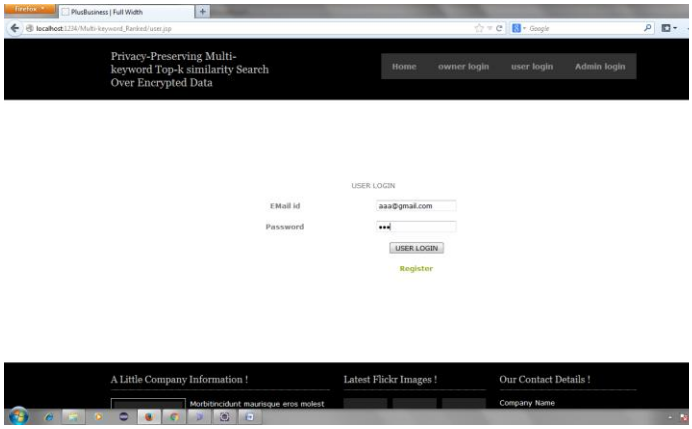
Owner home



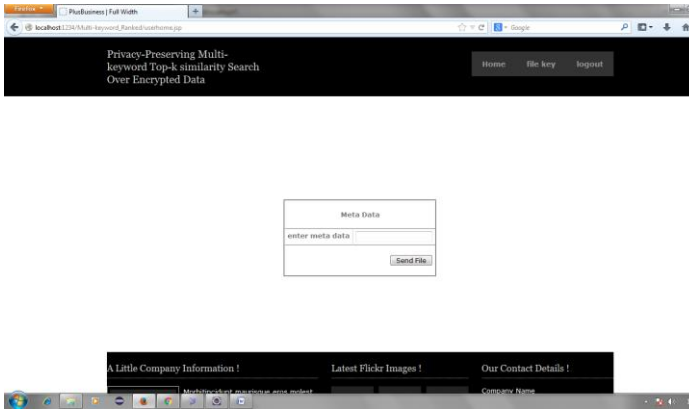
Click→File Uploading

Upload .txt file and click→send file, you will get meta data

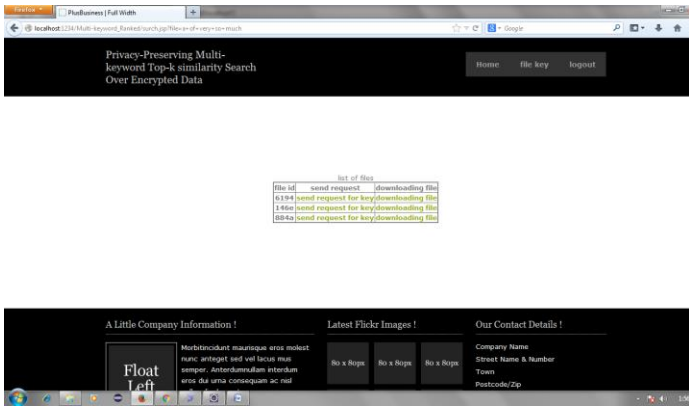
User login



Enter meta data of file→click send file

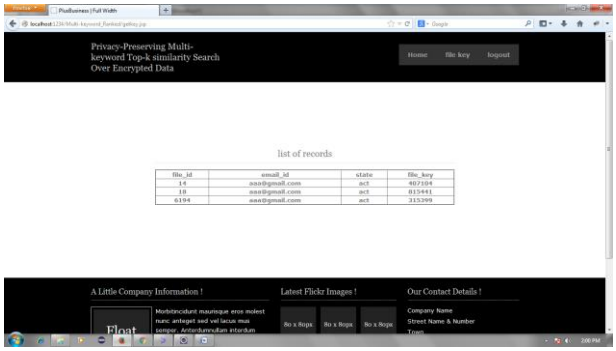


Click on send request for key

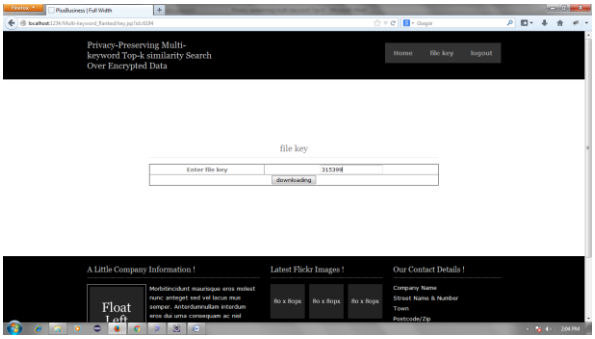


Go Back to Owner Page→Click Response→Click accept

Go to User Page→Click file key→Copy file_key



Click Home→enter Meta Data→Click Send File→Click Download→Enter Key→Download



Conclusion:

In this paper, we concentrate on rising the potency and the security of multi-keyword top-k similarity search over encrypted knowledge. At first, we tend to propose the random traversal algorithm which might deliver the goods that for 2 identical queries with completely different keys, the cloud server traverses completely different paths on the index, and also the knowledge user receives completely different results however with constant high level of question

accuracies in the unit of time. Then, so as to enhance the search efficiency, we tend to style the cluster multi-keyword top-k search scheme, that divides the lexicon into multiple teams and solely must store the top-ck documents of every word group once building index. Next, to guard the question unlikability, we tend to apply the random traversal algorithmic rule to get the RGMTS, which might increase the issue of cloud servers to conduct linkage attacks on 2 identical queries, and we may also tune the worth of E to form the amount of query unlikability versatile for knowledge house owners. Finally, the experimental results show that our ways are a lot of economical and safer than the progressive ways.

II. REFERENCES

[1]. J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, and R. Buyya, "Ensuring security and privacy preservation for cloud data services," ACM Computing Surveys, 2016.

[2]. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "A view of cloud computing," Communications of the ACM, vol. 53, no. 4, pp. 50–58, 2010.

[3]. R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proceedings of the 13th ACM Conference on Computer and Communications Security. ACM, 2006, pp. 79–88.

[4]. D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in Advances in CryptologyEurocrypt 2004. Springer, 2004, pp. 506–522.

[5]. Z. Ying, H. Li, J. Ma, J. Zhang, and J. Cui, "Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating," Sci China Inf Sci, vol. 59, no. 4, pp. 042 701:1–16, 2016.

[6]. D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Security and Privacy, 2000. SP 2000. Proceedings. 2000 IEEE Symposium on, 2000, pp. 44–55.

[7]. E.-J. Goh et al., "Secure indexes." IACR Cryptology ePrint Archive, vol. 2003, p. 216, 2003. [8] Y.-C. Chang and M. Mitzenmacher, "Privacy preserving keyword searches on remote encrypted data," in Applied Cryptography and Network Security. Springer, 2005, pp. 442–455. [9] Y. H. Hwang and P. J. Lee, "Public key encryption with conjunctive keyword search and its extension to a multi-user system," in Pairing-Based Cryptography–Pairing. Springer, 2007, pp. 2–22.

[8]. P. Golle, J. Staddon, and B. Waters, "Secure conjunctive keyword search over encrypted data," in Applied Cryptography and Network Security. Springer, 2004, pp. 31–45.

[9]. L. Ballard, S. Kamara, and F. Monrose, "Achieving efficient conjunctive keyword searches over encrypted data," in Information and Communications Security. Springer, 2005, pp. 414–426.

[10]. D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in Theory of cryptography. Springer, 2007, pp. 535–554.

[11]. E. Shen, E. Shi, and B. Waters, "Predicate privacy in encryption systems," in Theory of Cryptography. Springer, 2009, pp. 457–473.

[12]. 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 Proceedings of the 8th ACM SIGSAC Symposium on Information, ser. ASIA CCS '13. ACM, 2013, pp. 71–82.

[13]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," IEEE

Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014

Cite this article as :

V. Likhitha, Mr. A. Murali Mohan Kumar, "Improve Efficiency for Data Privacy Keywords Using Top-K Search Scheme", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 5, pp. 267-272, September-October 2022.
Journal URL : https://ijsrcseit.com/CSEIT228550