iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: VI Month of publication: June 2018

DOI: http://doi.org/10.22214/ijraset.2018.6113

Enhanced Confidentiality Multi-Keyword Top-k Identical Search Over Encrypted Data

Osman Hamid[1] and Md Ateeq Ur Rahman[2],
[1]Research Scholar, Dept. of Computer Science & Engineering, SCET, Hyderabad, India
[2]Professor and Head, Dept. of Computer Science & Engineering, SCET, Hyderabad, India

Abstract: Cloud computing provides people and enterprises huge computing power and climbable storage capacities to support a spread of massive information applications in domains like health care and research, thus a lot of and a lot of information homeowners area unit 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 concerning privacy considerations if the documents area unit discharged or shared to part untrusted third-parties in cloud. A sensible and wide used technique for information privacy preservation is to cipher information 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 tend to investigate the multi-keyword top-k search downside for large encoding against privacy breaches, ANd commit to establish an economical and secure answer to the current downside. Specifically, for the privacy concern of question information, we tend to construct a special tree-based index structure and style a random traversal rule, that makes even a similar question to provide completely different visiting methods on the index, and might conjointly maintain the accuracy of queries unchanged underneath stronger privacy.

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

I. INTRODUCTION

Cloud computing has emerged as a unquiet trend in each IT industries and analysis communities recently, its salient characteristics like high quantifiability and pay-as you- go fashion have enabled cloud shoppers to get the powerful computing resources as services per their actual necessities, such cloud users don't have any longer have to be compelled to worry regarding the wasting on computing resources and therefore the complexness on hardware platform management .

Nowadays, a lot of and a lot of firms and people from an oversized range of huge information applications have source their information and deploy their services into cloud servers for straightforward information management, economical process} and question processing tasks. however once the businesses and people get pleasure from these benefits in cloud computing, they conjointly have to be compelled to take the privacy concern of the outsourced information under consideration. as a result of information sets in several applications typically contain sensitive data like e-mails, electronic health records and money dealing records, once {the information|the info|the information} owner outsourcing such sensitive data to the cloud servers that area unit thought-about to be partly sure, the info may be simply accessed and analyzed by cloud service suppliers lawlessly.

Since the analysis of those information sets might offer profound insights into variety of key areas in society (such as e-research, healthcare, medical and government services), therefore information house owners want effective, ascendible and privacy-preserving services before emotional their information to the cloud. encoding has been wide used for information privacy preservation in information sharing eventualities, it refers to mathematical calculation and algorithmic theme that remodel plaintext into cyphertext, that could be a non-readable type to unauthorized parties. a spread of knowledge cryptography models are projected and that they area unit accustomed write in code the info before outsourcing to the cloud servers. However, applying these approaches for encoding typically cause tremendous price in terms of knowledge utility, that makes ancient processing ways that area unit designed for plaintext information now not work overflow encrypted information.

The keyword-based search is such one wide used information operator in several info and knowledge retrieval applications, and its ancient process ways can't be directly applied to encrypted information. Therefore, the way to method such queries over encrypted information and at a similar time guarantee information privacy becomes a hot analysis topic. fortuitously, several methodologies supported searchable cryptography are studied. let's say, modify the one keyword search, and works support the multi-keyword mathematician search.

However, the one keyword search isn't sensible enough to support advanced queries and therefore the mathematician search is false since it causes high communication price. Therefore, more moderen works like concentrate on the multikeyword hierarchical search, that is a lot of sensible in pay-asyou- go cloud paradigm. however most of those ways cannot meet the high search potency and therefore the robust information security at the same time, particularly once applying them to huge encoding poses nice quantifiability and potency challenges.

Motivated by this, during this paper, we have a tendency to concentrate on a special sort of multi-keyword hierarchical search, particularly the multikeyword top-k search, that has been a awfully well-liked info operator in several vital applications, and solely must come the k documents with the very best relevancy scores. For supporting multi-keyword search, we have a tendency to introduce the vector house model that represents documents and queries as vectors. so as to support top-k search, the relevancy scores between documents and queries ought to be calculated, therefore, the TF_IDF (term frequency eight inverse document frequency) model is introduced as a coefficient rule to reason the relevancy scores for ranking functions. additionally, to enhance the question potency for higher user experiences, we have a tendency to propose a gaggle multi-keyword top- k search theme (GMTS), that is predicated on partition and supports top-k similarity search over encrypted information.

In this theme, the info owner divides the keywords within the lexicon (suppose that the lexicon contains all the keywords that would be extracted from all documents) into multiple teams and establishes a searchable index for every cluster. On the opposite facet, to higher management the dimensions of indexes, we have a tendency to adopt champion lists into our theme, wherever the index of a keyword cluster solely stores the top-ck documents of the corresponding keyword (the top-ck documents of a keyword represent the c eight k documents that have the very best relevancy scores to the present keyword, wherever c could be a positive integer).

Furthermore, we have a tendency to propose a random traversal algorithmic rule (RTRA) to strengthen the info security, wherever the info owner builds a binary tree as searchable index and assigns a random switch to every node, that the information user will assign a random key to every question. Therefore, the info user will modification the results and visiting ways of queries by victimization completely different keys, that maintains high accuracy of queries. Finally, we have a tendency to mix the GMTS and therefore the RTRA along into AN economical and secure answer to our projected drawback.

Our contributions may be summarized as follows: eight we have a tendency to initial propose the random traversal algorithmic rule that makes the cloud server every which way traverse on index and returns completely different results for a similar question, and within the in the meantime, it maintains the accuracy of queries unchanged for higher security. eight supported the random traversal algorithmic rule, we have a tendency to gift one each economical and secure searchable cryptography theme, which might support top-k similarity search over encrypted information. during this theme, the info owner will management the amount of question unlinkability while not sacrificing accuracy. eight Our experimental results show that our ways area unit a lot of economical than the progressive ways and might higher defend information privacy. Especially, our projected technique has sensible quantifiability performance once managing massive information sets.

II. RELATED WORKS

With the fast development of cloud computing, additional and additional enterprises/individuals square measure getting down to source native knowledge to the cloud servers. However, beneath open networks and not totally sure cloud environments, they face huge security and privacy risks (e.g., knowledge leak or revelation, knowledge corruption or loss, and user privacy breach) once outsourcing their knowledge to a public cloud or exploitation their outsourced knowledge. Recently, many studies were conducted to deal with these risks, and a series of solutions were planned to alter knowledge and privacy protection in untrusted cloud environments. to totally perceive the advances and see the analysis trends of this space, this survey summarizes and analyzes the progressive protection technologies. we have a tendency to 1st gift security threats associated needs of an outsourcing knowledge service to a cloud, and follow that with a high-level summary of the corresponding security technologies. we have a tendency to then linger over existing protection solutions to realize secure, dependable, and privacy-assured cloud knowledge services as well as knowledge search, knowledge computation, knowledge sharing, knowledge storage, and knowledge access. Finally, we have a tendency to propose open challenges and potential analysis directions in every class of solutions.

Cloud computing, a replacement readying and delivery model of computing resources, allows convenient network access to a virtualized pool of remote resources [Armbrust et al. 2010]. Its advantages embody on-demand self-service, unlimited resource pooling, broad network access, dynamic quantifiability, service-measured valuation, and alleviation of management risks [Xiao and Xiao 2013]. With of these advantages, cloud computing motivates individual and enterprise users to advisedly source their native knowledge to remote servers hosted by a Cloud Server supplier (CSP), that is, knowledge storage outsourcing. consistent with the info usage forecast by Gartner in 2012 [Verma 2012], users would store over a 3rd of their knowledge to the cloud by 2016. Besides knowledge storage service, users any expect to amass additional connected services from the cloud, similar to knowledge search, knowledge computation, knowledge sharing, and knowledge access. These cloud knowledge services1 might simply facilitate users avoid massive capital outlays and operational overheads for getting devices and managing them. A cloud is brought up as associate untrusted cloud atmosphere once its resources and services square measure open for public use and communication is performed over a nontrusted network. Generally, a public cloud (e.g., Amazon AWS, Microsoft Azure, and Google Cloud Platform) isn't totally sure by users. Therefore, though the advantages of this new cloud service paradigm square measure tremendous, serious security risks and privacy challenges square measure raised beneath untrusted cloud environments. significantly, once knowledge house owners source their knowledge to a public cloud, they'll lose tight management of the info as in their native storage systems. Curious cloud directors and unauthorized users might deliberately access the outsourced knowledge and procure the sensitive info for varied motivations. The investigation report of Verizon in 2015 [Verizon 2015] indicates that notorious knowledge breach incidents occurred from time to time in recent years. Moreover, knowledge corruption or loss might additionally happen in cloud servers attributable to failures incurred by improper configuration, computer code bugs, hardware errors, and power failures [Ko et al. 2013]. to avoid wasting space for storing and minimize prices, greedy CSPs might discard the info that square measure ne'er or seldom accessed by users, which can impact knowledge retrievability for users. The report from the Cloud Security Alliance (CSA) [CSA 2013] shows that the info security issues square measure among the highest threats within the cloud. Table I lists the protection threats to cloud computing known by CSA and Verizon [CSA 2013; KO et al. 2013; Verizon 2015]. in addition, once users use cloud knowledge services, privacy breaches typically occur thanks to undesirable interference from internal and external adversaries; let's say, a CSP will guess a user is unwell by perceptive his or her access to sure medical knowledge. Thus, it are often seen that cloud knowledge services square measure in and of itself not secure from the point of view of cloud users. If there aren't any effective security and privacy protection measures, it might be laborious to believe that cloud users can delegate a CSP to manage their knowledge solely supported price reductions and versatile services. To alleviate these considerations and consequently prompt the widespread readying of knowledge outsourcing services, leading cloud service suppliers (e.g., Amazon, Google, and Microsoft) developed completely different security measures to forestall knowledge exposure and amerciable access. let's say, they use AES-256 to encipher user knowledge at rest and leverage resource-based or user-based access policies to enforce knowledge access management in use [Amazon 2015a; Google 2015; Amazon 2015b]. However, these approaches cannot support versatile and ascendible knowledge sharing, privacy-assured knowledge search, and secure computation. except removing the storage management on the user facet, getting convenient knowledge utilization services within the cloud is precisely what users wish. Aiming at any addressing these problems, novel science primitives and varied security protection proposals for cloud knowledge services are given recently. they will be generally classified into four categories: confidentiality-assured cloud knowledge service, owner-controlled cloud knowledge sharing, integrity-guaranteed cloud knowledge storage, and privacy-preserving cloud knowledge access. additional specifically, searchable encoding and homomorphic encoding techniques square measure planned to enforce secure knowledge search and knowledge computation, respectively; selective encoding and attribute-based encoding techniques square measure introduced to realize approved access and secure knowledge sharing; obvious knowledge possession and proof-of-retrievability techniques square measure given to make sure knowledge ne plus ultra and retrievability; and privacy preservation is enabled to shield multiple dimensions of personal info (e.g., access pattern, question privacy, and identity information) once users access the info keep within the cloud. Note that we have a tendency to separate privacy preservation from knowledge confidentiality protection although privacy protection most likely are often achieved by sure knowledge confidentiality techniques. rather than pure knowledge protection, we have a tendency to concentrate on specific privacy protection goals in privacy preservation. during this article, we have a tendency to establish 2 broad classes of security threats to cloud knowledge services, that is, threats from external attackers and threats from internal participants, and gift four important security needs, that is, knowledge confidentiality, knowledge integrity, knowledge access managementlability, and privacy preservability. Following the wants, we offer an outline of the info and privacy protection solutions during a high level, which can provide readers an overview of protection technologies. particularly, we have a tendency to gift recent analysis advances of confidentiality protection, access control, integrity guarantee, and privacy preservation for secure cloud knowledge services. From this survey, a beginner or nonspecialist will simply follow this space to be told the issues. Moreover, we have a tendency to discuss some open challenges that require to be any explored, that provides future analysis directions for researchers during this space.

A. *Existing System*

Data encryption has been wide used for knowledge privacy preservation in knowledge sharing situations, it refers to mathematical calculation and algorithmic theme that rework plaintext into cypher text, that may be a non-readable type to unauthorized parties.A variety of knowledge coding models are planned and that they are wont to inscribe the information before outsourcing to the cloud servers.

However, applying these approaches for encoding sometimes cause tremendous value in terms of knowledge utility, that makes ancient processing strategies that are designed for plaintext knowledge not work brim over encrypted knowledge.The keyword-based search is such one wide used knowledge operator in several info and knowledge retrieval applications, and its ancient process strategies can-not be directly applied to encrypted knowledge

B. Disadvantages

The quality on hardware platform management.The data may be simply accessed and analyzed by cloud service suppliers illicitly.

III. PROPOSED SYSTEM

We first propose the random traversal algorithmic rule that makes the cloud server arbitrarily traverse on index and returns totally different results for constant question, and within the in the meantime, it maintains the accuracy of queries unchanged for higher security.Based on the random traversal algorithmic rule, we tend to gift one each economical and secure searchable encoding theme, which might support top-k similarity search over encrypted knowledge. during this theme, the info owner will management the extent of question unlinkability while not sacrificing accuracy.Our experimental results show that our strategies area unit a lot of economical than the progressive strategies and might higher defend knowledge privacy. Especially, our projected methodology has sensible measurability performance once managing massive knowledge sets. To improve the question potency for higher user experiences. we specialise in rising the potency and also the security of multi-keyword top-k similarity search

IV. SYSTEM ARCHITECTURE

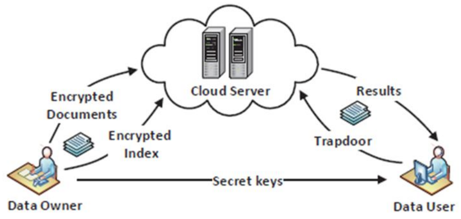

Fig. 1. The system model of searching over outsourced encrypted data.

Figure 1: System Architecture of the Proposed System

A. System Model

As shown in Fig. 1, the system model we tend to thought of during this paper contains 3 parts: the info owner, the info user and also the cloud server. the info owner uploads document assortment D to the cloud server, however this assortment might contain sensitive data. to guard information privacy, the info owner must encode D before outsourcing it to the cloud server. moreover, so as to modify the cloud server to method question expeditiously over the encrypted document assortment C, the info owner constructs AN encrypted searchable index id est domestically. Finally, the info owner outsources each the encrypted document assortment C and also the encrypted searchable index id est to cloud, and shares the key key of trapdoor generation and document decoding to licensed information users with secure channels. once the info user needs to go looking with a question , s/he generates the trapdoor T for this question first of all by query encoding, and so submits the trapdoor to cloud server for question process. when receiving T, the cloud server calculates the connection scores between trapdoor T and also the documents in index id est, and returns k documents with the best scores to the info user. Note that, the search management is outside the scope of our paper. Therefore, just like works [16], [2], [7], [3], [32], we tend to assume information users square measure trustworthy entities and also the trapdoors square measure generated by information users themselves.

B. Threat Model

In this paper, we tend to treat information owner and information user as trustworthy entities, however cloud server is taken into account to be ”honest-butcurious” as adopted in most works on secure cloud information search. The server is honest because it runs the programs and algorithms properly, it's curious since the cloud service suppliers will simply access and analyze the encrypted information, and even record queries to be told extra data. supported the data which may be learned by cloud sever, we tend to take into account 2 threat models as [15], [31].

Known Ciphertext Model. This threat model corresponds to the ciphertext-only attack, because the cloud server solely is aware of the encrypted document assortment C, encrypted searchable index id est and trapdoor T.

Known Background Model. Compared to the renowned ciphertext model, this model is additional stronger, because the cloud server here not solely is aware of the ciphertext of document assortment, searchable index and question, it's imagined to produce other information like data point data regarding the document assortment, which is able to expose additional information to cloud. as an instance, once the cloud server understands the normalized TF distributions of sure keywords, it will determine these keywords by examination the normalized TF distributions.

V. CONCLUSION

In this paper, we specialize in rising the potency and also the security of multi-keyword top-k similarity search over encrypted knowledge. At first, we propose the random traversal formula which might come through that for 2 identical queries with totally different keys, the cloud server traverses totally different ways on the index, and also the knowledge user receives totally different results however with identical high level of question accuracies within the mean solar time. Then, so as to enhance the search potency, we have a tendency to style the cluster multi-keyword top-k search theme, that divides the lexicon into multiple teams and solely must store the top-ck documents of every word cluster once building index. Next, to guard the question unlinkability, we have a tendency to apply the random traversal formula to urge the RGMTS, which might increase the issue of cloud servers to conduct linkage attacks on 2 identical queries, and that we may tune the worth of E to create the amount of question unlinkability versatile for knowledge homeowners. Finally, the experimental results show that our strategies are additional economical and safer than the progressive strategies.

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 Cryptology- Eurocrypt 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

[10] 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.