ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Multi Keyword Ranked Search Over Encrypted Cloud Data

[1]Abinaya Pasupathi, [2]G.Prabhakaran,[3]T.Rajasekaran

[1]Student, [2,3]Assistant Professor Department of Computer Science and Engineering Saveetha School of Engineering Saveetha University Chennai, India.

ARTICLE INFO

Article history:
Received 10 March 2015
Received 20 in revised form
March 2015
Accepted 25 March 2015
Available online 10 April 2015

Keywords:
Encryption, Cloud data, searching, ranked search, secure encryption.

ABSTRACT

With the advent of cloud computing, knowledgehouse owners are intended to source their complicatedknowledge management systems from native sites to business public cloud for nice flexibility and economic stashes. Excepting forprotectiveknowledge privacy, sensitive knowledge has to be encrypted before outsourcing, which obsoletes ancientknowledge utilization supported plaintext keyword search. Thus, sanctioningAssociate in Nursing encrypted cloud knowledge search service is of predominate importance. Considering the bigvarietyof information users and documents in cloud, it's crucial for the search service to allow multi-keyword questionand supply result similarity rankingto meet the effective knowledge retrieval want. Connected works on searchable codingspecialize in single keyword search or mathematician keyword search, and seldom differentiate the search outcomes. In this paper, for the primary time, we tend tooutline and solve the difficultproblem of privacy-preserving multi-keyword stratified search over encrypted cloud knowledge (MRSE), and establish a group of strict privacy requirements for such a secure cloud knowledge utilization systemto become a reality. Among varied multi-keyword linguistics, we select the economical principle of "coordinate matching", i.e., as numerous matches as achievable, to capture the similarity between search question and knowledge documents and more use "inner product similarity" toquantitatively formalize such principle for similarity activity. In this paper, proposalof a basic MRSE themeexploitation secure real number computation, and soconsiderably improve it to meet totally different privacy needs in 2 levels of threat models. Thoroughgoing analysis investigation privacy and potencyguarantees of plannedstructures is given, and experiments on the real-world dataset more show planned schemes so introduce low overhead on computation and communication.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: Abinaya Pasupathi, G.Prabhakaran, T.Rajasekaran., Multi Keyword Ranked Search Over Encrypted Cloud Data. ***Aust. J. Basic & Appl. Sci.,*** *9(11): 600-604, 2015*

INTRODUCTION

Cloud computing is that the long unreal vision of computing as a utility, wherever cloud customers will remotely store their knowledge into the cloud thusonget pleasure from the on-demand prime quality applications and services from a shared pool of configurable computing resources (Vaquero, L.M., *et al*., 2009). Its nice flexibility and economic savings are motivating eachpeople and enterprises to source their nativecomplicatedknowledge management system into the cloud, particularlyoncethe infomade by them that require to be hold on and utilized is chop-chop increasing. to safeguardknowledge privacy and combat unsought accesses in cloud and on the far side, sensitive knowledge, e.g., mails, personal health records, pic albums, tax documents, money transactions, etc., couldought to be encrypted by knowledgehouse owners before outsourcing to industrial public cloud (Kamara, S. and K. Lauter, 2010); this, however, obsoletes the standardknowledge utilization service supported plaintext keyword search. The trivial answer of downloading all the info and decrypting domestically is clearly impractical, because ofthe massivequantity of information measureprice in cloud scale schemes. Moreover, other than eliminating the inherent storage management, storing knowledge into the cloud serves no purpose unless they'll be simply searched and employed. Thus, discovering privacy-preserving and effective search service over encrypted cloud knowledge is of predominant importance. Considering the doubtless large number sizable quantity} of on-demand knowledge users and large amount of outsourced knowledge documents in cloud, this drawbackis especiallydifficultbecause it is verytoughto fulfilladditionallythe wants of performance, system utilization and quantifyiability. On the one hand, to achieve the effective knowledgerecoverywould like,

Corresponding Author: Abinaya Pasupathi, Student, Assistant Professor Department of Computer Science and Engineering Saveetha School of Engineering Saveetha University Chennai, India.
E-mail: abipasu5530@gmail.com

great amount of documents mandate cloud server to perform result relevancy ranking, rather thanrecurringuniformoutcome. Such hierarchal search system licensesknowledge users to search outthe foremost relevant data quickly, instead of burdensomely sorting through each match within the content assortment (Singhal, A., 2001). Hierarchal search can even elegantly eliminate extra network traffic by causing back solelythe foremost relevant knowledge that is very fascinating within the "pay-as-you- use cloud pattern. For privacy shield, such ranking operation, however, mustn't leak any keyword connecteddata. On the opposite hand, to enhance search result accuracy in addition as enhance user looking outexpertise, it'sadditionally crucial for such ranking system to support multiple keywords search, as single keyword search usually yields way too rough result. As a typicalfollowdesignated by today's internet search engines knowledge users could tend to producea collection of keywords rather thanonly 1because the indicator of their search interest to retrieve the foremost relevant knowledge. And every keyword within the search request is in a positionto assistslender down the search result additional. The "Coordinate matching" (Witten, I.H., *et al*., 1999), i.e., as several matches as achievable, is Associate in Nursingeconomical principle among such multi-keyword linguistics to refine the result relevancy, and has been extensiveemployed in the plaintext data retrieval (IR) community. However, a way to apply it within the encrypted cloud knowledge search system remains an awfullydifficult task thanks to inherent security and privacy hindrances, together withvaried strict necessities like knowledge privacy, index privacy, keyword privacy, and lots of others within the literature, searchable secret writing (Song, D., *et al*., 2000; Goh, E.-J., 2003; Chang, Y.-C. and M. Mitzenmacher, 2005; Curtmola, R., *et al*., 2006; Boneh, D., *et al*., 2004; Bellare, M., *et al*., 2007; Abdalla, M., *et al*., 2008; Li, J., *et al*., 2010; Boneh, D., *et al*., 2007) may be auseful technique that treats encrypted knowledge as documents and permits a user to firmly search over it through single keyword and retrieve documents of interest. However, direct application of those approaches to deploy secure giant scale cloud knowledge utilization system wouldn't be essentiallyappropriate, as they're developed as crypto primitives and can't accommodate such high service-level necessities like system utilization, user viewing outexpertise, and straightforwarddata discovery in mind.

Although some recent stylesareprojected to support mathematician keyword search (Golle, P., J. Staddon and B. Waters, 2004; Ballard, L., *et al*., 2005; Boneh, D. and B. Waters, 2007; Brinkman, R., 2007; Hwang, Y. and P. Lee, 2007; Katz, J., *et al*., 2008; Lewko, A., *et al*., 2010; Shen, E., *et al*., 2009) as an endeavorto complement the search tractability, they are still not capable ofoffer users with acceptable result ranking practicality (see section VI). Our early work (Wang, C., *et al*., 2010) has been awake to this downside, and resolves the safehierarchal search over encrypted knowledge with support of solely single keyword questions. Howevera way tostylesubordinate degreeeconomical encrypted knowledge search mechanism that supports multi-keyword linguisticswhile not privacy breaches still remains associate degreedifficult open downside. During this paper, for the primary time andwe have a trend tooutline and solve the matter of multi-keyword hierarchal search over encrypted cloud knowledge (MRSE) whereasconserving strict system-wise privacy in cloud computing paradigm. Among numerous multi-keyword linguistics, we decide the economical principle of "coordinate matching", i.e., as several matches as doable, to capture the similarity between search question and knowledge documents. Specifically, we have a tendency to use "inner product similarity" (Witten, I.H., *et al*., 1999), i.e., the amount of question keywords showingin a very document, to quantitatively judge the similarity of that document to the search question in "coordinate matching" principle. Throughout index construction, every document is related to a binary vector as a sub-index whereverevery bit represents whether or not corresponding keyword is contained within the document. The search questionis additionallydelineated as a binary vector whereverevery bit suggests thatwhether or not corresponding keyword seemsduring this search request, therefore the similarity can bespecifically measured by real number of question vector with knowledge vector. However, directly outsourcing knowledge vector or question vector can violate index privacy or search privacy. To fulfill the challenge of supporting such multi-keyword linguisticswhile not privacy breaches, we have a tendency to propose a basic MRSE themeexploitation secure real numbercomputation, whichis customized from a secure technique (Witten, I.H., *et al*., 1999), and so improve it step by step to realizenumerous privacy necessities in 2 levels of threat models. Our contributions area unit summarized as follows, 1) For the primary time, we have a tendency to explore the matter of multi-keyword hierarchal search over encrypted cloud knowledge, and establish a group of strict privacy necessities for such a secure cloud knowledge utilization system to become a reality. 2) We have a tendency to propose 2 MRSE schemes following the principle of "coordinate matching" whereas meeting completely different privacy necessities in 2 levels of threat models. 3) Thorough analysis work privacy and potency guarantees of projectedstructures is given, and experimentations on the real-world dataset additional show projected schemes so introduce low overhead on computation and communication.

System Architecture:

Consider a cloud knowledge hosting service involving three completely dissimilar entities such as knowledge owner, user, and cloud servers. Knowledge owner includes anassortmentof information documents F to be outsourced to cloud server within the encrypted type C. To modify the lookingability over C for effective knowledge utilization, data owner, before outsourcing, can1st build associate encrypted searchable index I from F, sosourceeach the index I and therefore the encrypted document assortment C to cloud server. To go looking the document assortment for t given keywords anda licensed user obtains a corresponding trapdoor T through search management mechanisms (Curtmola, R., *et al*., 2006). Upon acceptance T from knowledge users, cloud server is accountableto go looking the index I and come back the corresponding set of encrypted forms. To enhance document recovery accuracy, search result must to be hierarchic by cloud server reliable with some ranking criteria (e.g., coordinate matching, as are introduced shortly). Moreover, to scale backcommuniquéprice, knowledge user could send subordinateoptional range k at the side of the trapdoor T in order that cloud server solely sends back top- k documents that area unit most relevant to the search question. Finally, the access management mechanism is utilized to manage decipherment capabilities given to users. B.

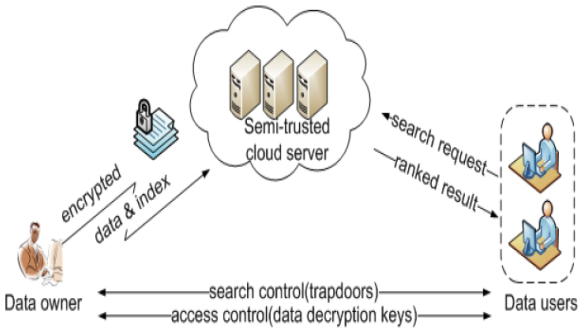

Fig. 1: System architecture

Threat Model Cloud server is taken into account as "honest-but-curious" in our model, that is in step withthe foremostconnected works on searchable coding. Specifically, cloud server acts in associate "honest" fashion and properly follows the selected protocol specification. However, it's "curious" to infer and analyze knowledge (including index) in its storage and message flows received throughout the protocol thuson learn extrainfo. Supported what info cloud server is aware of; we have a tendency totake into account2 levels of threat models as follows. Famous Cipher text Model during this model, cloud server is meant to solelyapprehend encrypted dataset C and searchable index I, each of thatarea unit outsourced from knowledge owner. Known Background ModelIn this stronger model, cloud server is meant to possess some backgrounds on the dataset, likethe topic and its connectedapplied mathinfo, furthermore to what may be accessed in famous cipher text model. As associate instance of potential attacks during this case, cloud server may utilize document frequency or keyword frequency (Zerr, S., *et al*., 2008) to spot keywords within thequestion.

Design Goals:

To changehierarchiclook for effective utilization of out-sourced cloud knowledgebeneath the said model, our system styleought toat the same timeaccomplish security and performance guarantees as follows.

1. Multi-keyword hierarchic Search:

To design search schemes which permit multi-keyword questionand supply result similarity ranking for effective knowledge retrieval, rather than returning uniform results.

2. Privacy-Preserving:

To prevent cloud server from learning extrainfo from dataset and index, and to satisfy privacy needsper section III-B.

3. Efficiency:

Above goals going onpracticality and privacy must to be achieved with low communication and computation overhead

Related Work:

Single Keyword Searchable codingancient single keyword searchable coding schemes (Song, D., *et al*., 2000; Goh, E.-J., 2003; Chang, Y.-C. and M. Mitzenmacher, 2005; Curtmola, R., *et al*., 2006; Boneh, D., *et al*., 2004; Bellare, M., *et al*., 2007; Abdalla, M., *et al*., 2008; Li, J., *et al*., 2010) Boneh, D., *et al*., 2007; Wang, C., *et al*., 2010) typically build Associate in Nursing encrypted searchable index specified its content is hidden to the server unless it's given applicable trapdoors generated via secret key(s) (Kamara, S. and K. Lauter, 2010). Its first studied by Song *et al*. (Song, D., *et al*., 2000) within thecruciate key setting and enhancements and advanced security definitions areagreed in Goh (Goh, E.-J., 2003) and Chang *et al*. (2005) and Curtmola *et al*. (2006). Our initial work (Wang, C., *et al*., 2010)

solves safehierarchal keyword search that utilizes keyword frequency to rank results rather than returning differentiated results. Conversely, it solelyprovisions single keyword search. Within the public key setting, Boneh *et al*. (2004) gift the first searchable coding construction, wherever anyone with public key will write to the informationhold on server howeversolelyapproved users with non-public key will search. Public key solutions aretypicallyterribly computationally big-ticketbut. Furthermore, the keyword privacy couldn't be dwindlingwithin the public key setting since server mayinscribe any keyword with public key then use the received trapdoor to judge this cipher text. Boolean Keyword Searchable codingto counterpoint search functionalities, conjunctive keyword search (Golle, P., *et al*., 2004; Ballard, L., *et al*., 2005; Boneh, D. and B. Waters, 2007) over encrypted knowledge has been estimated. These schemes sustainmassivedirectly above caused by their basic primitives, like computation value by additive map (Boneh, D. and B. Waters, 2007), or communication worth by secret distribution, e.g. (Ballard, L., *et al*., 2005). As aadditional general search approach, predicate coding schemes (Katz, J., *et al*., 2008; Lewko, A., *et al*., 2010; Shen, E., *et al*., 2009) are recently projected to support each conjunctive and divisional search. Conjunctive keyword search returns "all-or-nothing", which implies it exclusivelyyields those documents during which all the keywords specified by the search question appear; divisional keyword search returns differentiated results, which implies it yieldseach document that contains a set of the specific keywords, even only one keywords of interest. In short, none of prevailingBoolean keywords searchable coding schemes support multiple keywords hierarchal search over encrypted cloud knowledgewhereasprotective privacy as we tend to propose to explore during this paper. Dot product a query in predicate coding solely predicates whether or not2 vectors ar orthogonal or not, i.e., the dot productworth is hid except once it equals zero. While not providing the potentialto checkhid inner product, predicate codingisn't qualified for actinghierarchal search. Moreover, most of those schemes aredesigned upon the big-ticketanalysis of pairing operations on elliptic curves. Such inefficiency disadvantage conjointly limits their sensible performance once deployed in cloud. On a unique front, the analysis on top-k retrieval (Zerr, S., *et al*., 2009) in information community is additionally loosely connected to our downside.

Conclusion:

In this paper, for the first time we tend to define and solve the matter of multi-keyword hierarchal search over encrypted cloud knowledge, and establish a spread of privacy requirements. Among diverse multi-keyword linguistics, we designate the efficient principle of "coordinate matching", i.e., as numerous matches as achievable, to effectively capture similarity between question keywords and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity activity. For meeting the challenge of supporting multi-keyword linguisticswhile not privacy openings, we tend to first suggest a basic MRSE themevictimizationsafedot product computation, and significantly improve it to realize privacy needs in 2 levels of threat models. A thorough analysis investigation privacy and efficiency guarantee of projected schemes is given, and experiments on the real-world dataset show our projected schemes introduce low overhead on each computation and communication. As our future work, we are going to explore supporting alternative multi-keyword linguistics (e.g. weighted query) over encrypted knowledge, integrity checkered of ordering in search result and privacy guarantees in stronger threat model.

REFERENCES

Abdalla, M., M. Bellare, D. Catalano, E. Kiltz, T. Kohno, T. Lange,J. Malone-Lee, G. Neven, P. Paillier, and H. Shi, 2008. "Searchable encryptionrevisited: Consistency properties, relation to anonymous ibe, andextensions," J. Cryptol., 21(3): 350-391.

Ballard, L., S. Kamara and F. Monrose, 2005. "Achieving efficient conjunctivekeyword searches over encrypted data," in Proc. of ICICS.

Bellare, M., A. Boldyreva and A. ONeill, 2007. "Deterministic and efficientlysearchable encryption," in Proc. of CRYPTO.

Boneh, D. and B. Waters, 2007. "Conjunctive, subset, and range queries onencrypted data," in Proc. of TCC, pp: 535–554.

Boneh, D., E. Kushilevitz, R. Ostrovsky, and W.E.S. III, 2007. "Public keyencryption that allows pir queries," in Proc. of CRYPTO.

Boneh, D., G.D. Crescenzo, R. Ostrovsky and G. Persiano, 2004. "Public keyencryption with keyword search," in Proc. of EUROCRYPT.

Brinkman, R., 2007. "Searching in encrypted data," in University of Twente,PhD thesis.

Chang, Y.-C. and M. Mitzenmacher, 2005. "Privacy preserving keywordsearches on remote encrypted data," in Proc. of ACNS.

Cohen, W.W., "Enron email dataset," http://www.cs.cmu.edu/

Curtmola, R., J.A. Garay, S. Kamara and R. Ostrovsky, 2006. "Searchablesymmetric encryption: improved definitions and efficient constructions,"in Proc. of ACM CCS.

Goh, E.-J., 2003. "Secure indexes," Cryptology ePrint Archive, 2003,http://eprint.iacr.org/2003/216.

Golle, P., J. Staddon and B. Waters, 2004. "Secure conjunctive keyword searchover encrypted data," in Proc. of ACNS, pp: 31-45.

Hwang, Y. and P. Lee, 2007. "Public key encryption with conjunctive keywordsearch and its extension to a multi-user system," in Pairing.

Ishai, Y., E. Kushilevitz, R. Ostrovsky and A. Sahai, 2009. "Cryptography fromanonymity," in Proc. of FOCS, pp: 239-248.

Kamara, S. and K. Lauter, 2010. "Cryptographic cloud storage," in RLCPS,January 2010, LNCS. Springer, Heidelberg.

Katz, J., A. Sahai and B. Waters, 2008. "Predicate encryption supportingdisjunctions, polynomial equations, and inner products," in Proc. OfEUROCRYPT.

Lewko, A., T. Okamoto, A. Sahai, K. Takashima and B. Waters, 2010. "Fully secure functional encryption: Attribute-based encryption and(hierarchical) inner product encryption," in Proc. of EUROCRYPT.

Li, J., Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, 2010. "Fuzzy keywordsearch over encrypted data in cloud computing," in Proc. of IEEEINFOCOM'10 Mini-Conference, San Diego, CA, USA.

Shen, E., E. Shi and B. Waters, 2009. "Predicate privacy in encryptionsystems," in Proc. of TCC.

Singhal, A., 2001. "Modern information retrieval: A brief overview," IEEEData Engineering Bulletin, 24(4): 35-43.

Song, D., D. Wagner and A. Perrig, 2000. "Practical techniques for searcheson encrypted data," in Proc. of S&P.

Vaquero, L.M., L. Rodero-Merino, J. Caceres and M. Lindner, 2009. "A breakin the clouds: towards a cloud definition," ACM SIGCOMM Comput.Commun. Rev., 39(1): 50-55.

Wang, C., N. Cao, J. Li, K. Ren and W. Lou, 2010. "Secure ranked keywordsearch over encrypted cloud data," in Proc. of ICDCS'10.

Witten, I.H., A. Moffat and T.C. Bell, 1999. "Managing gigabytes: Compressingand indexing documents and images," Morgan Kaufmann Publishing,San Francisco.

Wong, W.K., D.W. Cheung, B. Kao and N. Mamoulis, 2009. "Secure knncomputation on encrypted databases," in Proc. of SIGMOD.

Zerr, S., D. Olmedilla, W. Nejdl and W. Siberski, 2009. "Zerber+r: Top-kretrieval from a confidential index," in Proc. of EDBT.

Zerr, S., E. Demidova, D. Olmedilla, W. Nejdl, M. Winslett and S. Mitra, 2008. "Zerber: r-confidential indexing for distributed documents," in Proc. OfEDBT, pp: 287-298.