Trends in Engineering and Technology (NCTET-2K17)
International Journal of Advanced Engineering Research and Science (IJAERS) *Special Issue-3*
https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

A Systematic Approach for Multi-Keyword Ranked Search over Encrypted Data by Using User Interest Model

V.Deepika[1], D.Shobha Rani[2]

[1]PG scholar, Department of CS, Sri Venkateswara Engineering College for Women (SVEW), TIRUPATI - 517501
Email: deepika.vuppalapati@gmail.com
[2] Associate Professor, Department of CSE, Sri Venkateswara Engineering College for Women (SVEW), TIRUPATI
Email: shobharani.d@svcolleges.edu.in

Abstract:-In distributed computing, searchable encryption conspire over outsourced information is a hot research field. Nonetheless, generally existing deals with encoded seek over outsourced cloud information take after the model of "one size fits all" and overlook customized look expectation. Additionally, the greater part of them bolster just correct watchword look, which enormously influences information ease of use and client encounter. So how to outline a searchable encryption conspire that backings customized seek and enhances client look encounter remains an extremely difficult assignment. In this paper, interestingly, we consider and tackle the issues of customized multi-watchword positioned seek over encoded information (PRSE) while protecting security in distributed computing. With the assistance of semantic philosophy WordNet, we manufacture a client intrigue demonstrate for singular client by investigating the client's pursuit history, and embrace a scoring instrument to express client premium cleverly. To address the restrictions of the model of "one size fit all" and watchword correct inquiry, we propose two PRSE plans for various pursuit expectations. Broad investigations on true dataset approve our examination and demonstrate that our proposed arrangement is exceptionally productive and viable.

Keywords-Cloud security, outsourcing security, personalized search, user interest model.

I. INTRODUCTION

Distributed computing has accomplished awesome improvement both in scholastic and industry groups as it gives financial and advantageous administration. What's more, now an ever increasing number of organizations and clients are wanting to transfer their information onto general society mists. Be that as it may, information put away in the cloud may experience the ill effects of malevolent use by cloud specialist co-ops since information proprietors have at no time in the future direct control over information. Considering information protection and security, it is a prescribed practice for information proprietors to encode information before transferring onto the cloud. Despite the fact that it shields information security from illicit utilize both from untrusted cloud specialist co-ops and outside clients, it makes information use more troublesome since numerous systems in light of plaintext are no longer appropriate to cipher text. In this way, investigating an effective scan procedure for encoded information is to a great degree dire. A mainstream approach to look over scrambled information is searchable encryption and numerous helpful plans have been advanced under various applications. Be that as it may, these searchable encryption plans in view of catchphrase never again completely fulfill the new test and clients' expanding needs, particularly showed in the accompanying two angles.

In those plans, the cloud will restore all documents that match the client's inquiry, which may bring about a colossal utilization of system data transmission. In addition, it will cost client much time and numerous assets to channel his genuine fascinating ones among a vast amount of returned documents. In the commonsense application, diverse clients may discover distinctive things significant in light of various significance or needs of inquiry terms, demonstrating the need of customized hunt, which considers individual catchphrase inclination or watchword need.

The other one is that the vast majority of these plans bolster just correct catchphrase seek. That implies the returned result is just identified with the client's information. At the point when the client inquiries some unprecedented terms, it is conceivable that just a couple coordinated outcomes are returned and the client may be not happy with the returned comes about. Moreover,

https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

most clients are ordinarily untrained and easygoing, they may not input the unequivocal inquiry terms that precisely coordinate their actual inquiry goals. Therefore, the returned result is certainly not what clients truly need. In such cases, clients need to recover more outcomes as comparable as conceivable to inquiry terms. To take care of this issue, the regular system utilized as a part of IR is inquiry expansion, which can stretch out unique question terms as indicated by a few standards before submitting look ask. Likewise, question augmentation stays to be balanced and enhanced to meet accessible encryption in distributed computing.

1.1 PROBLEM STATEMENT

An accessible encryption plot with support of both customized positioning and question expansion is the issue that we attempt to handle in this paper. In this paper, interestingly, we examine and take care of the issue of customized multi-watchword positioned seek over encoded information (PRSE) while protecting security in the distributed computing. In PRSE, with the assistance of semantic cosmology WordNet, client intrigue display for individual client is worked by examining the client's inquiry history. What's more, we receive a scoring component to express client intrigue cleverly by computing the similitude score between various sorts of related words and the watchword. We propose an essential outline of PRSE, and after that give two PRSE plans in light of secure internal item keeping in mind the end goal to meet distinctive hunt expectations.

1.2 RELATED WORK

One-Time Verifier-Based Encrypted Key Exchange provides strong security arguments to support the EKE -like protocols being standardized by the IEEE P1363.2 Standard working group (namely the PPK series). We have reached this aim by slightly modifying the original AuthA protocol (the two encryption primitives are instantiated using separate mask generation functions but derived from a unique shared password) to be able to achieve the security notion of forward-secrecy in a provably-secure way. Our result is a slight departure from previously known results on EKE -like structures since the security of AuthA is now based on the Gap Diffie-Hellman problem. Moreover, we have extended AuthA into a One-time Password-authentication and Key exchange (OPKeyX) technology which allows a user to securely log into his account using a remote un-trusted computer and limits the damages of corruption of the server.

Authenticated key exchange secure against dictionary attacks provides Classic cryptographic protocols based on user-chosen keys allow an attacker to mount password-guessing attacks. A combination of asymmetric (public-key) and symmetric (secret-key) cryptography that allow two parties sharing a common password to exchange confidential and authenticated information over an insecure network is introduced. In particular, a protocol relying on the counter-intuitive motion of using a secret key to encrypt a public key is presented. Such protocols are secure against active attacks, and have the property that the password is protected against offline dictionary attacks.

The Decision Diffie-Hellman assumption (ddh) is a gold mine. It enables one to construct efficient cryptographic systems with strong security properties. In this paper we survey the recent applications of DDH as well as known results regarding its security. We describe some open problems in this area.

Searchable Encryption

The principal useful accessible encryption plot in symmetric setting is proposed by Song et al. in which each word is scrambled freely under a twolayer development and the client needs to experience the entire report to look a specific catchphrase. And after that some security definitions and numerous enhancements or developments have been proposed by Goh Chang et aland Curtmola et al.. Boneh et al. propose the primary open key-based accessible encryption plot, where anybody owning the private key can look information things scrambled by the general population key. As of late, D. Money et al. and E. Stefanov et al. examine the issue of data spillage in conjunctive inquiry and dynamic accessible symmetric encryption, individually. From that point forward, a ton of plans under various applications have been proposed. Among them, to address the spelling botch, import issue of fluffy catchphrase hunt is proposed by Li et al. and enhanced in. What's more, Wang et al. and Cao et al. do inquire about on secure positioning on single catchphrase and multi-watchword separately. From that point onward, Sun et al. enhance the productivity of multi-watchword look by embracing MDB-tree. In any case, a large portion of existing accessible encryption plans bolster just correct catchphrase seek, which influences information ease of use and client's understanding. Fu et al. propose a semantic catchphrase look conspire in view of stemming calculation, which enables clients to discover applicable records containing

https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

semantically close watchwords identified with the question word. Moreover, customized inquiry is additionally missed or disregarded. Shen et al. proposed a favored catchphrase look plot over scrambled information, however how to quantify watchword inclination is overlooked. The simulated technique for measuring catchphrase inclination is tedious and forces a weight on the client. Also, it neglects to consider diverse clients' hunt histories and in this manner has awesome irregularity.

Personalized Search

Customized seek goes for misusing client data to empower indexed lists better meet the individual client's inquiry goal. The general approach is to construct a client profile, which portrays the client's advantages or inclinations that can straightforwardly set by the client or gathered amid the look history for a period. What's more, the most difficult works in customized pursuit are: 1) how to manufacture a client profile demonstrate; 2) how to make utilization of the client profile to enhance the hunt. The client profile model is regularly based upon an arrangement of watchword vectors or classes of catchphrase vectors. Because of the nonappearance of interrelation of watchwords in the catchphrase based portrayal display, a few analysts make utilization of an arrangement of ideas gotten from predefined metaphysics or reference cosmology to express the client profile demonstrate. It is outstanding that metaphysics based client profile model is better than other portrayal models as it can abuse semantic learning. Take note of that hunt personalization is accomplished by incorporating the client profile in the question reformulation handle inquiry record coordinating customized report order or archive re-positioning. The momentum web indexes, for example, GOOGLE, has propelled customized seek, where the client can show his interests or inclinations expressly, or inclinations might be naturally accumulated. Be that as it may, they have not been generally embraced yet since clients stress over the spillage of client protection. Besides, a large portion of existing customized seek plans are inapplicable to ciphertext.

1.3 EXISTING SYSTEM

To address the impediment of the model of "one size fits all" in most existing accessible encryption plans. Because of the way that a large portion of existing accessible encryption plans bolster just correct watchword seek, it is conceivable that the client just gets a couple comes about by questioning a few terms . a large portion of existing accessible encryption plans bolster just correct watchword look, which influences information ease of use and client's understanding. Be that as it may, they have not been generally received yet since clients stress over the spillage of client protection. Additionally, a large portion of existing customized seek plans are inapplicable to ciphertext.

1.4 PROPOSED SYSTEM

A preferred keyword search scheme over encrypted data, but the artificial manner of measuring keyword preference has great randomness and fails to consider different users' search histories. Keyword Weight Keywords are practical tools to summarize document content.

Keyword Priority Through the well-trained user interest model, we can get the access frequency of each keyword. The higher the access frequency of a keyword is, the more important the keyword is from viewpoint of the user. The importance of the keyword usually means its rank priority.

Relevance Score It is used to measure the score of the query to a document. We can divide the whole relevance score into many sub-scores to represent the connection of file to the keywords in the query.

Secure Inner Product When calculating the relevance score of the document to the query, we should use two vectors: the document vector and the query vector. However, it is not advisable to directly outsource two vectors onto the cloud at the risk of leaking index privacy and query privacy.

II SYSTEM ARCHITECTURE

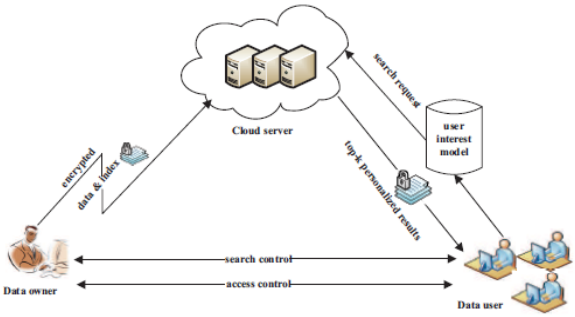

Fig. 1: Architecture of the search over encrypted cloud data

A complete system model in cloud computing should involve three different entities: the data owner, the data user and the cloud server. Different from the previous work there exists a user interest model stored in

https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

the user side. The user interest model is built upon the user's long-term search history. It records access frequency of both query keywords and their related keywords with the help of WordNet. Different access frequency of keywords, as keyword priority can reflect their different importance in viewpoint of the data user. To search for files of interest, the data user should firstly produce a search request. And then query reformulation that achieves keyword priority of query terms will be carried out through the user interest model. At last, the encrypted search query through search control mechanism, e.g., broadcast encryption, will be sent to the cloud. Upon receiving the search request from the authorized user, the cloud server will conduct designated search operation over the index and send back relevant encrypted document.

III MODULES

1. User Interest Model Construction
2. Searchable Encryption
3. Personalized Search

MODULE DESCRIPTION

1. USER INTEREST MODEL CONSTRUCTION

So as to assess the development of client intrigue show, we arbitrarily select three clients, break down their noteworthy records and concentrate their question terms. The time taken a toll among three clients as for various size of notable records. We can see the time cost ascends with number of memorable records as for a client. Because of assorted qualities of notable records among the clients, their time taken a toll in a similar size of noteworthy records changes generally. Additionally, we record the capacity overhead of client intrigue demonstrate with various size of inquiry terms.

The time cost to create an inquiry for the most part relies on upon the quantity of catchphrases in the word reference, since the basic primary operation or time consuming operation in every one of the plans is question encryption. So the time cost will turn out to be vast as expanding the quantity of watchwords in the word reference.

2. SEARCHABLE ENCRYPTION

The main functional accessible encryption plot in symmetric setting is proposed by Song et al. in which each word is encoded autonomously under a twolayer development and the client needs to experience the entire report to look a specific watchword. And afterward some security definitions and numerous enhancements or developments have been proposed by Goh , Chang et al.and Curtmola et al. Boneh et al. propose the main open key-based accessible encryption plot, where anybody owning the private key can look information things encoded by the general population key. As of late, D. Money et al. what's more, E. Stefanov et al. talk about the issue of data spillage in conjunctive inquiry and dynamic accessible symmetric encryption, individually. From that point onward, a ton of plans under various applications have been proposed. Among them, to address the spelling botch, import issue of fluffy catchphrase pursuit is positioning on single watchword and multi-catchphrase individually. From that point onward, Sun et al. enhance the productivity of multi-watchword look by receiving MDB-tree. Be that as it may, the majority of existing accessible encryption plans bolster just correct catchphrase seek, which influences information convenience and client's involvement. Fu et al. propose a semantic catchphrase seek conspire in light of stemming calculation, which enables clients to discover applicable records containing semantically close watchwords identified with the question word. Besides, customized hunt is likewise missed or overlooked. Shen et al. proposed a favored catchphrase seek conspire over scrambled information, however how to quantify watchword inclination is overlooked. The counterfeit technique for measuring catchphrase inclination is tedious and forces a weight on the client. Additionally, it neglects to consider diverse clients' inquiry histories and in this manner has extraordinary arbitrariness. proposed by Li et al. also, enhanced in research on secure

3. PERSONALIZED SEARCH

Customized look goes for misusing client data to empower list items better meet the individual client's hunt goal. The general approach is to construct a client profile, which portrays the client's advantages or inclinations that can specifically set by the client or gathered amid the look history for a period. Furthermore, the most difficult works in customized inquiry are:

1) How to manufacture a client profile display
2) How to make utilization of the client profile to enhance the hunt.

The client profile model is regularly based upon an arrangement of catchphrase vectors or classes of watchword vectors. Because of the nonattendance of interrelation of watchwords in the catchphrase based portrayal show, a few scientists make utilization of an

https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

arrangement of ideas gotten from predefined philosophy or reference cosmology to express the client profile display. It is outstanding that philosophy based client profile model is better than other portrayal models as it can abuse semantic information. Take note of that inquiry personalization is accomplished by coordinating the client profile in the question reformulation prepare, inquiry record coordinating, customized report order or archive re-positioning. The ebb and flow web crawlers, for example, GOOGLE, has propelled customized seek, where the client can show his interests or inclinations expressly, or inclinations might be naturally assembled. Be that as it may, they have not been broadly embraced yet since clients stress over the spillage of client protection. Also, large portions of existing customized look plans are inapplicable to cipher text.

IV PERFORMANCE EVALUATION

I mainly focus on the precision of our scheme. The precision here means whether users attain the desired documents according to queries or not. In our scheme, we also need balance precision and privacy. To reduce the loss of precision, balance strategy of precision and privacy by introducing of dummy keywords in our scheme. In theory, It's highly possible to return all the query results related to the users' recent search results because of the construction of users' interesting model (keywords in users' search history have high-priority). That brings great convenience to users who major on a certain subject recently and has a great impact on users' work.

To verify the precision of our scheme, we conduct a survey that 20 users are randomly invited to make a test for our system. As might be expected, more than three fourths are satisfied with the returned results by our system. This also proves the effectiveness of our system. We also conduct a survey on our documents set to perform top-k (we set k=20) search, and all the 20 returned documents accord with our query. The first one of the returned results is what we need. However, in the documents set, there are 100 documents related to our query. The recall rate is 0.2, but the most related documents are returned.

V CONCLUSION

We tackle the problem on customized multi-keyword ranked ask over encrypted planet data. Considering the person search history, we construct a person pastime mannequin because individual user along the help concerning semantic philosophy WordNet. Through the model, we hold cooked automatic contrast regarding the key-word priority and solved the issue over the artificial method over measuring. Moreover, we advocate joining PRSE schemes in imitation of resolve two barriers (the model on "one bulk fit all" or key-word precise search) between nearly existing searchable encryption schemes. In addition, thoroughgoing privacy analysis then performance analysis demonstrates so our design is practicable.

VI REFERENCES

[1] E. Shi, J. Bethencourt, T. H. H. Chan, and D. Song, A. Perrig. "Multi-Dimensional Range Query over Encrypted Data," in Proc. Of IEEE Symposium on Security and Privacy (SP'07), 2007, pp. 350C364.

[2] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. j. Lou, "Fuzzy keyword search over encrypted data in cloud computing," in Proc. of IEEE INFOCOM'10 Mini-Conference, San Diego, CA, USA, March 2010.

[3] C. Liu, L. H. Zhu, L. Li, and Y. Tan, "Fuzzy Keyword Search on Encrypted Cloud Storage Data with Small Index," in Proc. of IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), 2011, pp. 269-273.

[4] C. Wang, N. Cao, J. Li, K. Ren, and W. J. Lou, "Secure Ranked Keyword Search over Encrypted Cloud Data," in Proc. of IEEE 30th International Conference on Distributed Computing Systems (ICDCS), 2010, pp. 253-262.

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

[6] M. Chuah, W. Hu, "Privacy-aware BedTree Based Solution for Fuzzy Multi-keyword Search over Encrypted Data," in Proc. of 31st International Conference on Distributed Computing Systems Workshops (ICDCSW), 2011, pp. 273-281.

[7] C. Wang, K. Ren, S. C. Yu, and K. M. R. Urs, "Achieving Usable and Privacy-assured Similarity Search over Outsourced Cloud Data," in Proc. of IEEE INFOCOM 2012, 2012, pp. 451-459.

[8] M. Islam, M. Kuzu, and M. Kantarcioglu, "Access pattern disclosure on searchable encryption: Ramification, attack and mitigation," In Proc.of NDSS'12, 2012.

[9] C. Liu, L. Zhu, M. Wang, and Y. Tan, "Search Pattern Leakage in Searchable Encryption: Attacks and New Constructions," Cryptology ePrint Archive, Report 2013/163, 2013, http://eprint.iacr.org/.

https://dx.doi.org/10.22161/ijaers/nctet.2017.14 *ISSN: 2349-6495(P) | 2456-1908(O)*

[10] Z. Shen, J. Shu, and W. Xue, "Preferred keyword search over encrypted data in cloud computing," In Proc. of 21st International Symposium on Quality of Service (IWQoS'13), 2013.
[11] W. K. Wong, D. W. Cheung, B. Kao, and N. Mamoulis, "Secure knn computation on encrypted databases," in Proc. of SIGMOD, 2009, pp. 139-152.
[12] X. Shen, B. Tan and C. Zhai, "Context-Sensitive Information Retrieval Using Implicit Feedback," in Proc. of SIGIR, 2005, pp. 43-50.
[13] W. Fan, M. D. Gordon, and P. Pathak, "Discovery of contextspecific ranking functions for effective information retrieval using genetic programming," IEEE Trans. on Knowl. and Data Eng., vol. 16, no. 4, pp. 523-527, 2004.
[14] Z. Ma, G. Pant, O. R. L. Sheng, "Interest-based personalized search", ACM Trans. Inf. Syst., vol. 25, no. 1, 2007.
[15] A. Sieg, B. Mobasher, and R. Burke, "Web search personalization with ontological user profiles," in Proc. of CIKM'07, 2007, pp. 525- 534.
[16] "Enron Email Dataset," https://www.cs.cmu.edu/_enron/, last accessed 11-20-2013.
[17]"Enron Dataset," http://www.isi.edu/_adibi/Enron/Enron.htm, last accessed 11-20-2013.
[18] "The Enron email dataset database schema and brief statistical report," http://www.isi.edu/_adibi/Enron/Enron Dataset Report. pdf, last accessed 11-20-2013.
[19] M. F. Porter, "An Algorithm for Suffix stripping", Automated Library and Information Systems, vol. 14, no. 3, pp. 130-137, 1980
[20] M. Ondrei˘cka and J. Pokorn´ y, "Extending fagin's algorithm for more users based on multidimensional b-tree," in Proc. of ADBIS, 2008, pp. 199-214.
[21] 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 Proc. of ACM ASIACCS, 2013, pp. 71-82.
[22] Z. Fu, J. Shu, X. Sun, and D. Zhang, "Semantic keyword search based on trie over encrypted cloud Data," in Proc. of SCC'14, 2014, pp. 59-62.
[23] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, "Highly-scalable searchable symmetric encryption with support for boolean queries," Cryptology ePrint Archive, Report 2013/169, 2013, http://eprint.iacr.org/.
[24] E. Stefanov, C. Papamanthou, and E. Shi, "Practical dynamic searchable encryption with small leakage," Cryptology ePrint Archive, Report 2013/832, 2013, http://eprint.iacr.org/.
[25] D. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. of IEEE Symposium on Security and Privacy'00, 2000.