A FAST NEAREST NEIGHBOR SEARCH SCHEME OVER OUTSOURCED ENCRYPTED CLOUD DATA

G. Sreelekha[1], Sk. Shayeer Hussain[2]

[1]Assistant Professor, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

[2]PG Scholar, Dept. of MCA, Audisankara College of Engineering and Technology (AUTONOMOUS), Gudur, AP, India.

ABSTRACT

Cloud computing has generated much interest in the research community in recent years for its many advantages, but has also raise security and privacy concerns. The storage and access of confidential documents have been identified as one of the central problems in the area. In particular, many researchers investigated solutions to search over encrypted documents stored on remote cloud servers. While many schemes have been proposed to perform conjunctive keyword search, less attention has been noted on more specialized searching techniques. In this paper, we present a phrase search technique based on Bloom filters that is significantly faster than existing solutions, with similar or better storage and communication cost. Our technique uses a series of n-gram filters to support the functionality. The scheme exhibits a trade-off between storage and false positive rate, and is adaptable to defend against inclusion-relation attacks. A design approach based on an application's target false positive rate is also described.

I. INTRODUCTION

As organizations and individuals adopt cloud technologies, many have become aware of the serious concerns regarding security and privacy of accessing personal and confidential information over the Internet. In particular, the recent and continuing data breaches highlight the need for more secure cloud storage systems. While it is generally agreed that encryption is necessary, cloud providers often perform the encryption and maintain the private keys instead of the data owners. That is, the cloud can read any data it desired, providing no privacy to its users. The storage of private keys and encrypted data by the cloud provider is also problematic in case of data breach. Hence, researchers have actively been exploring solutions for secure storage on private and public clouds where private keys remain in the hands of data owners.

Boneh et al. [1] proposed one of the earliest works on keyword searching. Their scheme uses public key encryption to allow keywords to be searchable without revealing data content. Waters et al. [2] investigated the problem for searching over encrypted audit logs. Many of the early works focused on single keyword searches. Recently, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords [3], [4]. Other interesting problems, such as the ranking of search results [5], [6], [7] and searching with keywords that might contain errors [8], [9] termed fuzzy keyword search, have also been considered. The ability to search for phrases was also recently investigated [10], [11], [12], [13]. Some [14] have examined the security of the proposed solutions and, where flaws were found, solutions were proposed [15].

In this paper, we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data. We begin by presenting the communication framework in section 2 and various backgrounds including related works in section 3. Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches. Therefore, we describe both the basic conjunctive keyword Search algorithm and the basic phrase search algorithm in section 4 along with design techniques in section 4.3. Performance analysis and experimental results are included in section 5 and 6.

II. LITERATURE SURVEY

Inpaper[1]conferredaphrasesearchthemesupported Bloom filter that's considerably quickerthanexistingapproaches,requiringsolelyone sphericalofcommunicationandBloomfilterverifications. The answer addresses the high processvaluenotedinbyreformulatingphrase.Theirapproach is also the primary to effectively permitphrasesearchtorunseverallywhilenotfirstplay actingaconjunctivekeywordsearchtospotcandidate documents. The technique of constructingaBloomfilterindexintroducedallowsquickverification of Bloom filters in the same manner ascompartmentalization[6].Strengths:scalebackstorage value and provide security within the sort offalse positivesandadaptthethemetodefendagainst inclusionrelationattacks.Weakness:Theverification speed is less and fewer communicationvalue.

In paper [7] they proposed a viable way to deal withtake care of the issue of equivalent word based multiwatchwordpositionedseekoverencodedcloudinformation. The filed records can be refined whenaffirmedcloudcustomersinputthecomparableexpressions of the predefined catchphrases, not theright or cushy organizingwatchwords, due tothepossible proportionate word substitution and also hernonappearanceofrightfindingoutaboutthedata.Forthe first time they formalize and manage the issue ofsupportingefficientyetsecurityensuringpaddedlookforaccomplishingproductiveuseofremotelysetawayblendedinformationinCloudComputing.Strengths: Computationcomplexnessisgreatlyreduced and improves the potency of the server toretrieve the encryption information. Weakness:Theserver cannotgenerate trapdooritself.

In paper [8] they have format an induced approach togatherthebreakingpointefficientdelicatecatchphrase sets by mauling a significant wisdom onthecomparabilitymetricofprogressdivided.Incontext of the created padded watchword sets, theyhaveadditionallyproposedanefficientcushycatchphrase look design. Through cautious securityexamination, they show that our proposed strategy issecure and confirmation guarding, while effectivelyunderstandingtheobjectiveofcushycatchphraselook.

In paper [9] they proposed a multi-catchphrase lookplotinlightofWangetal'sconspire.Theyadditionally novel technique for watchword changesandpresentsthestemmingcalculation.Theirplandoesnotrequireapredefinedcatchphrasesetandthusempowers efficient file refresh. In this paper, theyexaminetheissueofmulti-catchphrasecushionedsituated investigate mixed cloud data. They propose amulticatchphrasesoftsituatedlookforplaninperspective of Wang et al's plot. Weakness-Theseschemes aim solely to protect the keyword set of asinglequestion,whereastherelationsbetweendifferentqueriesdon'tseemtobestudied.

Inpaper [10]additionalstudiedthematterofsearchableencoding, that solves the perplexity of maintainingthe confidentiality of knowledge and also the abilityfor a consumer to search. They have introduced themodelofphrasesearchwithsymmetricencodinganditssecuritydefinition,andthenproposeaconstruction and its security proof. They have proved that theirschemeachievesnon adaptivesecurity.Strengths: It achieves non accommodative security.Weakness:Itdoesn'tmeetthestandardsof adaptivesecurity.

In paper [6] conferred a phrase search theme basedmostlyonBloomfilterthatachieveseighttimeslowerstoragevalueintheirexperimentthantheprevailingsolutionswhereasexhibitingsimilarorhighercom municationandprocessrequirements.The planned resolution provides the basic rankingcapability,maybecustom-madetonon-keywordsearch and is appropriate against inclusion-relationattacks [11].Strengths:Theflexibilitytolookovertheencryptedinformationandprovidesthebasicrankingcapability,maybecustom-madetonon-keywordsearch and is appropriate against inclusion relationattack. Weakness:Totally different split values mayleakinformationonthedocumentcontent.

Inpaper[14]theyaskedabouttheissueofarticulationanalyzemixeddataandproposedadynamicmulti-phraseorchestratedscanforoverencoded data with symmetric open encryption. Notthesameaspriorwork,ourarrangementenablesdatacustomerstolookthroughacoupleofarticulationsinademandrequest,andthedataproprietorcanvitalizethe outsourced data at less cost. Remembering thetrueobjectivetoranktherundownthings,theyfoundthecentralityscoresinsidetheTFIDFappearonclient side. It conceivably keeps up a key divisionfrom the spillage of significance scores. The novelsynopsisassociateswithdatacustomerstolookencoded data successfully.

In paper [15] they propose another MRSE structurewhich beats each and every one of the bits of theKNN-SE based MRSE systems. Specifically, theirnewsystemdoesnotrequireapredefinedwatchwordsetandsponsorshipscatchphrasesinsubjectivetongues,isamulti-customerstructurewhichreinforcesflexiblerequestguaranteeingandtime-controlled foreswearing, and it achieves better datasecurity attestation since even the cloud server can'ttell which records are the best k occurs obviouslyreturned to a data customer. They proposedmulti-catchphrase rank open encryption which vanquishesevery last one of the defects of the KNN-SE basedMRSEframeworks.

Inpaper [17]theyproposedmulti-keywordranksearchable encryption which conquers every one oftheimperfectionsoftheKNN-SEbasedMRSEframeworks. The framework permits flexible huntapprovalandtimecontrolleddisavowal.Theydemonstratedthesecurityoftheframeworkanddirectedbroad PCre-enactmentstoshowitsefficiency.Strengths:Thesystemallowsflexiblesearch authorization and time-controlled revocation.Inpaper[16]theystickandhandletheissue ofsecuremulti-watchword top-k recovery over blended cloudinformation. They definedrespectability congruityandplanquality.Inlightoforderpreservingencryptionbafflinglyreleasescrappydata;they deviseaserversideplanningSSEmake.Theybythenpropose a two-round open encryption (TRSE) plotutilizingtheabsolutelyhomomorphicencryption, whichfulfillsthesecuritystraybitsofmultiwatchwordtoprecoveryovertheencodedcloudinformation.Bysecurityexamination,theyshowthatthe proposed plot ensures information inquire. Asshowedupbytheefficiencyappraisaloftheproposeddevise over declared dataset,wide testworks outunmistakably exhibit that our framework guaranteessensibleefficiency.

III. PROPOSED SYSTEM

In the proposed system, the system presents a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. The system also describes modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data.

In the proposed system, the system also presents a phrase search technique based on Bloom filters that is significantly faster than existing solutions with similar or better storage and communication cost. The proposed system technique uses a series of n-gram filters to support the functionality.

The scheme exhibits a trade-off between storage and false positive rate, and is adaptable to defend against inclusion-relation attacks. A design approach based on an application's target false positive rate is also described.

1. The Data retrieval is fast due to Conjunctive keyword search scheme.
2. The security is more on outsourced data due to Modified phrase search scheme against IR attacks.

Waters et al. [2] investigated the problem for searching over encrypted audit logs. Many of the early works focused on single keyword searches. Recently, researchers have proposed solutions on conjunctive keyword search, which involves multiple keywords [3], [4]. Other interesting problems, such as the ranking of search results [5], [6], [7] and searching with keywords that might contain errors [8], [9] termed fuzzy keyword search, have also been considered. The ability to search for phrases was also recently investigated [10], [11], [12], [13]. Some [14] have examined the security of the proposed solutions and, where flaws were found, solutions were proposed [15].

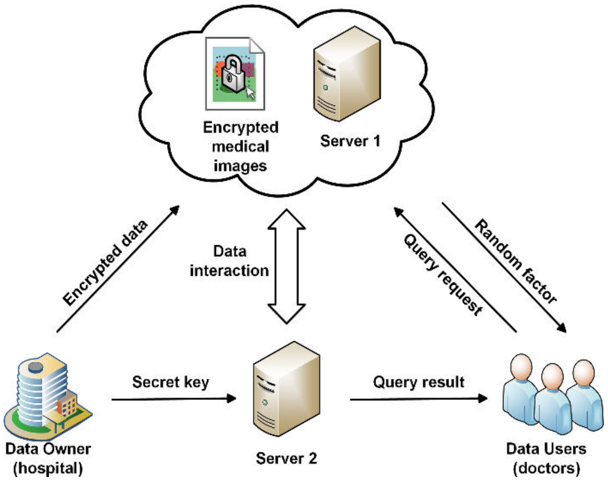

Fig: Architecture of the proposed methodology

In this paper, we present a phrase search scheme which achieves a much faster response time than existing solutions. The scheme is also scalable, where documents can easily be removed and added to the corpus. We also describe modifications to the scheme to lower storage cost at a small cost in response time and to defend against cloud providers with statistical knowledge on stored data. We begin by presenting the communication framework in section 2 and various backgrounds including related works in section 3. Although phrase searches are processed independently using our technique, they are typically a specialized function in a keyword search scheme, where the primary function is to provide conjunctive keyword searches. Therefore, we describe both the basic conjunctive keyword search algorithm and the basic phrase search algorithm in section 4 along with design techniques in section.

VI. CONCLUSION

In this paper, we presented a phrase search scheme based on Bloom filter that is significantly faster than existing approaches, requiring only a single round of communication and Bloom filter verifications. The solution addresses the high computational cost noted in [13] by reformulating phrase search as n-gram verification rather than a location search or a sequential chain verification. Unlike [10], [12],[13], our schemes consider only the existence of a phrase, omitting any information of its location. Unlike [11], our schemes do not require sequential verification, is parallelizable and has a practical storage requirement. Our approach is also the first to effectively allow phrase search to run independently without first performing a conjunctive keyword search to identify candidate documents. The technique of constructing a Bloom filter index introduced in section 4.2 enables fast verification of Bloom filters in the same manner as indexing. According to our experiment, it also achieves a lower storage cost than all existing solutions except [13], where a higher computational cost was exchanged in favor of lower storage. While exhibiting similar communication cost to leading existing solutions, the proposed solution can also be adjusted to achieve maximum speed or high speed with a reasonable storage cost depending on the application. An approach is also described to adapt the scheme to defend against inclusion-relation attacks. Various issues on security and efficiency, such as the effect of long phrases and precision rate, were also discussed to support our design choices.

V. REFERENCES

[1]D. Boneh, G. D. Crescenzo, R. Ostrovsky, and G. Persiano, “Public key encryption with keyword search,” in In proceedings of Eurocrypt, 2004, pp. 506–522.

[2] B. Waters, D. Balfanz, G. Durfee, and D. K. Smetters, “Building an encrypted and searchable

audit log," in Network and Distributed System Security Symposium, 2004.

[3] M. Ding, F. Gao, Z. Jin, and H. Zhang, "An efficient public key encryption with conjunctive keyword search scheme based on pairings," in IEEE International Conference on Network Infrastructure and Digital Content, 2012, pp. 526–530.

[4] F. Kerschbaum, "Secure conjunctive keyword searches for unstructured text," in International Conference on Network and System Security, 2011, pp. 285–289.

[5] C. Hu and P. Liu, "Public key encryption with ranked multi keyword search," in International Conference on Intelligent Networking and Collaborative Systems, 2013, pp. 109–113.

[6] Z. Fu, X. Sun, N. Linge, and L. Zhou, "Achieving effective cloud search services: multi-keyword ranked search over encrypted cloud data supporting synonym query," IEEE Transactions on Consumer Electronics, vol. 60, pp. 164–172, 2014.

[7] C. L. A. Clarke, G. V. Cormack, and E. A. Tudhope, "Relevance ranking for one to three term queries," Information Processing and Management: an International Journal, vol. 36, no. 2, pp. 291–311, Jan. 2000.

[8] H. Tuo and M. Wenping, "An effective fuzzy keyword search scheme in cloud computing," in International Conference on Intelligent Networking and Collaborative Systems, 2013, pp. 786–789.

[9] M. Zheng and H. Zhou, "An efficient attack on a fuzzy keyword search scheme over encrypted data," in International Conference on High Performance Computing and Communications and Embedded and Ubiquitous Computing, 2013, pp. 1647–1651.

[10] S. Zittrower and C. C. Zou, "Encrypted phrase searching in the cloud," in IEEE Global Communications Conference, 2012, pp. 764–770.

[11] Y. Tang, D. Gu, N. Ding, and H. Lu, "Phrase search over encrypted data with symmetric encryption scheme," in International Conference on Distributed Computing SystemsWorkshops, 2012, pp. 471–480.

[12] H. Poon and A. Miri, "An efficient conjunctive keyword and phrase search scheme for encrypted cloud storage systems," in IEEE International Conference on Cloud Computing, 2015.

[13] "A low storage phrase search scheme based on bloom filters for encrypted cloud services," to appear in IEEE International Conference on Cyber Security and Cloud Computing, 2015.

[14] V. Srikanth, et al. “Detection of Fake Currency Using Machine Learning Models.” Deleted Journal, no. 41, Dec. 2023, pp. 31–38. https://doi.org/10.55529/ijrise.41.31.38.

[15] V. Srikanth, et al. “A REVIEW ON MODELING AND PREDICTING OF CYBER HACKING BREACHES.” 25 Mar. 2023, pp. 300–305. http://ijte.uk/archive/2023/A-REVIEW-ON-MODELING-AND-PREDICTING-OF-CYBER-HACKING-BREACHES.pdf.

[16] V. Srikanth, “DETECTION OF PLAGIARISM USING ARTIFICIAL NEURAL NETWORKS.” 25 Mar. 2023, pp. 201–209. http://ijte.uk/archive/2023/DETECTION-OF-PLAGIARISM-USING-ARTIFICIAL-NEURAL-NETWORKS.pdf.

[17] V. Srikanth, “CHRONIC KIDNEY DISEASE PREDICTION USING MACHINELEARNINGALGORITHMS.” 25 January. 2023, pp. 106–122. http://ijte.uk/archive/2023/CHRONIC-KIDNEY-DISEASE-PREDICTION-USING-MACHINE-LEARNING-ALGORITHMS.pdf.

[18] Srikanth veldandi, et al. “View of Classification of SARS Cov-2 and Non-SARS Cov-2 Pneumonia Using CNN”. journal.hmjournals.com/index.php/JPDMHD/article/view/3406/2798.

[19] Srikanth veldandi, et al. “Improving Product Marketing by Predicting Early Reviewers on E-Commerce Websites.” Deleted Journal, no. 43, Apr. 2024, pp. 17–25. https://doi.org/10.55529/ijrise.43.17.25.

[20] Srikanth veldandi, et al.“Intelligents Traffic Light Controller for Ambulance.” Journal of Image Processing and Intelligent Remote Sensing, no. 34, July 2023, pp. 19–26. https://doi.org/10.55529/jipirs.34.19.26.

.AUTHOR’S PROFILE

G. SREELEKHAis currently working as Assistant Professor in Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.

SK. SHAYEER HUSSAIN is pursuing MCA from Audisankara College of Engineering and Technology (AUTONOMOUS), NH-5, Bypass Road, Gudur, Tirupati (Dt.), Andhra Pradesh, India.