The Efficient and Probabilistic Symmetric Searchable Encryption Scheme in Cloud

*D. Vijaya Durga[1], A.P.V.D.L. Kumar[2]
[1]M.Tech Scholar, Department of Computer Science & Engineering,
[2]Assosciate.Professor, Department of CSE, BVC Institute of Technology & Science,
Batlapalem, Amalapuram, AP, India.

Abstract—

Cloud have enormous storage space for storing immense size of data. Data owner redistribute their data substance on cloud server. Cloud server can have enormous storage space. In this paper, data client demand for data to the cloud server. Data owner produce encryption key for each datum client which mentioned for data files. Because of this, data stays secure and precise data may looked by semantic pursuit. Additionally the data spillage is limited because of the cloud storage framework. Calculation can be utilized for encryption system. Key age is performed by data administrator for protection safeguarding. Presently multi day's there will develop prominence of cloud computing, huge number of clients and data proprietors are spurred to re-appropriate their data to cloud servers for enormous comfort and decreased cost required for data the board. Be that as it may, significant data ought to be scrambled before redistributing for security necessities, which uses data use strategy like watchword based archive recovery. Accessible encryption is of extending energy for guaranteeing the data assurance in secure accessible circulated storage. In this paper, we explore the security of a remarkable cryptographic crude, specifically, open key encryption with watchword search which is incredibly significant in various uses of cloud storage.

Keywords: Cloud Computing, SAAS, MRSE, Cloud security, File data integrity, Public key encryption,.

I. Introduction

Cloud computing is the utilization of computing assets (equipment and programming) that are conveyed as a service over a system (ordinarily the Internet). The name originates from the basic utilization of a cloud molded image as a reflection for the unpredictable foundation it contains in framework graphs. Cloud computing depends remote services with a client's data, programming and calculation. Cloud computing comprises of equipment and programming assets made accessible on the Internet as oversaw outsider services. These services regularly give access to cutting edge programming applications and top of the line systems of server PCs. Cloud computing is a normal model in the developing patterns of Information Technology. It empowers consistent access to shared pool of configurable framework assets. It tends to be effectively dealt with insignificant exertion. It gives security and availability all the while. Cloud computing includes the utilization of computing assets to convey services over system. It depends remote services with client's data, programming and calculation. The data which was put away under an outsider framework was defenseless against robbery. Cloud storage redistributing is of expanding enthusiasm for late years for undertakings and associations to diminish the weight of keeping up enormous data. Truly, end clients may like to encode their re-appropriated data for security assurance as they may not by any means trust the cloud storage server. This makes sending of customary data usage service, for example, plaintext catchphrase search over literary data or question over database, a troublesome assignment. One of the ordinary arrangements is the accessible encryption which enables the client to look and recover the encoded data, and then save the data security. Tragically, notwithstanding being free from mystery key circulation, PEKS schemes experience the ill effects of an innate security issue with respect to the watchword protection, specifically (inside) disconnected Keyword Guessing Attack (KGA). In particular, given a trapdoor, the ill-disposed server can pick a speculating watchword from the catchphrase space and after that utilization the

Revised Manuscript received on August 12, 2019
*Corresponding Author
D. Vijaya Durga,
mail id- vijaydurga2011@gmail.com

watchword to create a PEKS figure content. The server at that point can test whether the speculating catchphrase is the one basic the trapdoor. This speculating then-testing methodology can be rehashed until the right catchphrase is found. As the catchphrase consistently could release some delicate data of the client data, it is hence of handy significance to beat this security risk for secure and accessible encoded data re-appropriating.

II. Related work

The creator in [9] Cryptographic Cloud Storage paper said that when the advantages of utilizing an open cloud foundation are clear, it presents critical security and protection dangers. Truth be told, it appears that the greatest snag to the reception of cloud storage (and cloud computing all in all) is worry over the confidentiality and honesty of data. In [9], an outline of the advantages of a cryptographic storage service, for instance, lessening the legitimate introduction of the two clients and cloud suppliers, and accomplishing administrative consistence is given. Other than this, cloud services that could be based over a cryptographic storage service, for example, secure reinforcements, authentic, wellbeing record frameworks, secure data trade and e-revelation is expressed quickly. In [10], answers for this issue under well-characterized security necessities are advertised. To get this, think about the issue: a client U needs to store his files in a scrambled structure on a remote document server S. Later the client U needs to effectively recover a portion of the encoded files containing explicit catchphrases, keeping the watchwords themselves mystery and not to jeopardize the security of the remotely put away files. For instance, a client might need to store old email messages scrambled on a server overseen by Yahoo or another huge seller, and later recover certain messages while going with a cell phone. The schemes are effective as no open key cryptosystem is included. To be sure, the methodology is autonomous of the encryption strategy picked for the remote files. They are gradual as well. In that, client U can submit new files which are secure against past inquiries yet at the same time accessible against future questions. From this, the primary subject taken is of putting away data remotely on other server and recovering that data from anyplace through portable, PC and so on. The creators in tells the significance of securing person's protection in cloud computing and gives some protection safeguarding advances utilized in cloud computing services. As protection is a significant issue for cloud computing, both as far as legitimate consistence and client trust and should be considered at each period of structure. Paper said that it is critical to consider while structuring cloud services, if these include the accumulation, handling or sharing of individual data. From this paper, primary topic taken is of saving security of data. This paper just portrays protection of data yet doesn't permit recorded pursuit just as doesn't shroud client's identity. Therefore, these two disadvantages are defeated in our proposed framework. The creator in this paper, recommended a calculation for unknown sharing of private data among N gatherings is created. This strategy is utilized iteratively to relegate these hubs ID numbers extending from 1 to N. This task is mysterious in that the characters got are obscure to different individuals from the gathering. In, existing and new calculations for allotting unknown IDs are analyzed as for exchange offs among correspondence and computational necessities. These new calculations are based over a protected aggregate data mining activity utilizing Newton's personalities and Sturm's hypothesis. The principle thought taken from this paper is of allocating mysterious ID to the client on the cloud. The creator in this paper propose a semantic multi-watchword positioned search conspire over the encoded cloud data, which at the same time meets a lot of exacting protection necessities. Right off the bat, we use the "Dormant Semantic Analysis" to uncover connection among terms and records. The connection between terms is naturally caught. Furthermore, our plan utilize secure "k-closest neighbor (kNN)" to accomplish secure hunt usefulness. The proposed plan could return the definite coordinating files, yet in addition the files including the terms inactive semantically related to the question catchphrase. Here the proposed framework in the interim backings dormant semantic inquiry and utilizations the vectors comprising of TF values as lists to records. These vectors establish a grid, from which we dissect the inert semantic relationship among terms and archives by LSA.

III. Methodology

Secure inquiry over scrambled information have been recently connected to cloud server. Wang et al. proposed secure inquiry conspire over scrambled cloud information. In accessible encryption, customers store information into encoded structure to the cloud server and watchword looking can be perform on ciphertext. Accessible encryption (SE) strategies [5], [6], [7], [8], can halfway satisfy the requirement for secure redistributed information search. Secure pursuit over encoded cloud information lessens the calculation and capacity cost. Secure positioned multi-catchphrase search, fluffy watchword search, similitude search every one of these quests are likewise performed on scrambled cloud information. Information client validation procedure, Different-key scrambled watchwords coordinating and protection saving positioned search of records techniques are utilized to tackle the issue of secure multi-catchphrase scan for various information proprietors and numerous information clients in distributed computing condition. At the point when enormous measure of information proprietors [3], [9] are included then they produce trapdoors all the while which influence the adaptability and convenience of hunt framework.

A. Information User Authentication Technique:

Information client verification procedure is utilized to keep framework from aggressors who claiming to be lawful information clients performing look. Ming Li [3] proposed fine-grained approval structure in which client gets search capacities under nearby confided in specialists (LTAs). Outsider examiners (TPA) used to validate information client before playing out any looking on cloud server [4]. Another system to give protection from assailants is client denial [3], [9], where information client can't play out any quests once he is disavowed.

B. Coordinating Different-Key Encrypted Keywords:

Early works for the most part just help single watchword search. Afterward, a few multi-catchphrase search plans were proposed [5], [6], [7], [8]. Information proprietor store information in encoded structure and information client create trapdoors [3], [4] to send inquiry demand in scrambled structure. Re-encryption of watchword file and trapdoors [9] used to build greater security from aggressors. Wenhai Sun [5], proposed tree-based list structure with the goal that down to earth search effectiveness is greatly improved than direct inquiry. Ning Cao [6], proposed organize coordinating which gives however many matches as would be prudent which catch the significance of information archives to the pursuit inquiry and inward item similitude to quantitatively assess such closeness measure. Zhihua Xia proposed a plan which supports dynamic update tasks like erasure of reports and inclusion of archives and treebased list structure and Greedy Depth first Search calculation use to give effective multi-catchphrase positioned search. Hongwei Li bolster confused rationale search by utilizing the blended AND, OR and NO tasks of catchphrases for commonsense and proficient multi-watchword search conspire. Proposed issue of customized multi-catchphrase positioned search over scrambled cloud information. A client intrigue model is worked for individual client with the assistance of semantic philosophy WordNet by utilizing client search history.

C. Protection Preserving Ranked Search:

In accessible symmetric encryption plans, because of enormous number of archives, query items ought to be recovered in a request for the significance with the looked through catchphrases. Scoring is the normal method to weight the pertinence of the records. TFIDF [4], [6], [7], [8] is notable strategy to register the pertinence score. Wong et al. [13] proposed a safe k-closest neighbor (kNN) plot which can secretly encode two vectors and register Euclidean separation of them [6], [8].

IV. Established Research Gaps Related To Privacy

One such difficulty arises with key word privacy. This is attributed to the tendency of users to prefer concealing their search from others including the cloud service provider whereby the most pressing concern is masked what they are searching. These usually constitutes the keywords that correspond to the trapdoor. The problem is that despite the fact that the trapdoor can be generated in a cryptographic way for the protection of the keywords of a search, there is a possibility that the cloud service provider, through statistical analysis of the search result, can come up with almost if not accurate estimates [1]. The second privacy concern in multi-keyword searches relates to trapdoor un-linkability. This concern draws from the fact that the trapdoor generation tool, as opposed to being randomized is deterministic. This is important because it prevents the relationships between trapdoor from being revealed to the cloud service provider. With a deterministic trapdoor generation, the cloud service provider is handed the advantage of accumulating frequencies of different queries that regard different keywords [1]. This, in essence, means that the privacy requirement of the keyword is undermined and thus it leads to privacy concerns. Thirdly, there is also a concern relating to access. The pattern of access refers to the sequence of search results within the ranked data, where every search result relates to a set of documents in a rank order [1]. If a query is linked to a set W, whereby its search result is expressed as FW and consists of the identification list of all documents by their relevance, the resulting access pattern can be generated in sequence. Despite a number of proposed encryption models in which the main framework is based on private information retrieval, the analyzed studies are not designed to conceal the pattern of access. In view of the aforementioned limitations and issues, a starting point would be to initially sort out the privacy-based requirements that align correctly with an innovative and efficient system. As such, the privacy requirements should at its very least, not require users to encrypt their data/keywords through other encryption methods while making a multi-keyword search. Additionally, a more appropriate model for privacy-preserving multi-keyword search could comprise of the following four modules as adopted from the study by Dhumal & Jhadav [3];

- Binary data generation.
- Data ciphering.
- Data user access control.
- Data user query.

V. Concurrency Control Protocols

In what follows, we briefly present the most prominent concurrency control protocols that can be used in cloud database.

Self-optimizing One Copy Serializability (SO- 1SR)

1SR is the strongest and well known correctness criterion for applications that are newly deployed in the cloud. It assures the serializable execution of concurrent transactions and a one copy view of the data. The most commonly used approaches to implement 1SR is to use lock based protocols such as strict two-phase locking (S2PL) for providing serializable transaction execution and two-phase commit (2PC) for synchronous updating all replicas.

Transaction model:

In a system providing 1SR, each transaction which writes to a data object must update all copies of the data object. In case of update transactions the replicated data increases the response time and thus decreases the overall scalability of the system. In order to exploit the merits of the cloud, it is essential to provide scalability, availability, low cost and strongly consistent data management. Under distributed systems, it is not possible to provide consistency and availability. The stronger consistency level decreases the availability and scalability. In cloud environments, the cost of guaranteeing a certain consistency level on top of replicated data is to be considered. Strong consistency is costly; on the other hand, weak consistency is cheaper, but may lead to high operational costs of compensating the effects of anomalies and access to stale data. The first generation cloud DBMS's provide on the weak consistency in order to provide maximum scalability and availability. It is sufficient for satisfying requirements related to consistency of simple cloud applications. However, more sophisticated like web shops, online stores and credit card services requires strong consistency levels. The advantages of cloud such as availability and scalability are not yet exploited by existing commercial and open source DBMS's which provide strong consistency. SO-1SR (self-optimizing 1SR) is a novel protocol for replicated data in a cloud that dynamically optimize all phases of transaction executions. System model of SO-1SR assumes that applications are built on the top of a cloud data environment.

Implementation:

The SO-1SR middleware should be present at each replica node. The transactions that are submitted by the client to the application servers are forwarded to the SO-1SR middleware for optimal execution. The SO-1SR is based on a fully replicated system and flat transaction model. Protocols like 2PC or Paxos are needed to provide strong consistency guarantees. The main goal of SO-1SR is to decrease latency by using dynamic optimization technique at different phases of transaction life cycle, not to replace protocols like 2PC or Paxos. .

Snapshot Isolation

The transactional guarantees of SI are weaker than 1SR, such that the database system can achieve increased concurrency by relaxing isolation requirements on transaction. In SI, the transaction attempting read is never blocked. The tradeoff between transaction isolation and performance is that higher degrees of transaction isolation assure fewer anomalies. Anomalies avoided by 1SR are also avoided in SI. Under SI, write skew anomaly is possible if two transactions concurrently update one or more common data item. For example, consider two transactions Tm and Tn. Transaction Tm reads data items p and q and then updates concurrently with other transaction Tn that reads data item p and q and then updates q. Here transaction Tm and Tn do not have a write-write conflict because none of the transaction updates a common data item. Different variations of SI exist for replicated systems like cloud which provide different consistency guarantees. In a lazily synchronized replicated database system; if two transactions Ts and Tv do not have a write–write conflict under SI, then their updates may be committed in the order Ts followed by Tv at a site S1 but in reverse order at another site S2 in which each site individually guarantees SI. In this case, consider a transaction Tk that reads x and y at site S1 and view database state from the commit of Ts will not view this same database state if it were to be executed on the database replica at site S2.But this kind of replica in consistency will not occur in a centralized database system that guarantees SI. SI was introduced by Berenson et al. SI is defined as; it does not allow dirty reads, dirty writes, non-repeatable reads, phantoms or lost updates. Write skew anomalies are possible in SI. By the definition of SI, when the transaction starts the system assigns a transaction Ta start timestamp called start (T). The database state seen by T is determined by start (T). The system can choose any time less than or equal to the actual start time of T to start (T). The update transactions made by Tl that commit after start (T) will not be visible to T. Only update transaction that commits before start (T) will be visible to T. Each transaction T is able to see its own updates are also a requirement in SI. Thus, if T updates a database item and reads that item, then T will see the updating even though the update occurred after the start (T).

Transaction model:

Commit timestamp, commit (T) is assigned to a transaction when a transaction is to commit. The time commit (T) is more recent than any other start or commit timestamp assigned to any transaction. If no other committed transaction Tk with lifespan [start (Tk), commit (Tk)] that overlaps with a T’s lifespan of [start (T), commit (T)] write data that T has also written then only T commits. Otherwise, to prevent lost updates T is getting aborted. This technique of preventing lost updates is called the first-committerwins (FCW) rule. Transaction inversions are possible in SI, i.e. for every pair of transactions T1 and T2, if T2 executes after T1 then T1 will view T1’s updates. This is because the actual start time of T2 can be larger than that of a start (T2). In particular, if T2 starts after T1 has finished, then T2 will see a database state that does not contain the effects of T1. In order to prevent these kinds of transaction inversions, strong SI is introduced. In the definition of strong SI (SSI), if for every pair of committed transactions Tp and Tq in transaction history TH such that Tp’s commit precedes the first operation of Tq, start (Tq) > commit (Tp) and it is SI then we can say that the transaction execution history TH is strong SI. 3.2.2. Implementation: The decentralized model of SI based transactions consists of some mechanisms such as: (a) Keeping a consistent, committed snapshot for reading (b) a global sequencer is used for arranging the transactions by allocating commit timestamps (c) detection of write-write anomalies in concurrent transactions and (d) atomically commit the updates and make them durable. In the model, each transaction goes through a sequence of phases during execution. The main phase is the active phase in which all read/write on data item is performed in this phase. The remaining phases are part of the commit of the transaction. Validation phase is required for detecting the conflicts among transactions that are executed concurrently.

Session Consistency Session

Consistency is considered to be the minimum consistency level in a distributed environment that does not result in complexities for application developers. Under Session Consistency, the application will not see its own updates and may get inconsistent data from successive accesses. The key idea is that, all data does not need the same level of consistency. There is a term called consistency rationing i.e. the data is divided into three categories A, B, C and each type of data is treated differently depending on the consistency level provided. The category A contains data in which consistency violations may result in large penalty costs. The category B includes data where the consistency requirements change over time. Category C comprises data in which inconsistency is acceptable. Session consistency considers data under category C. C category is always a preferred category for placing data in the cloud database [14]. By considering a transaction cost and response time the session consistency is very cheap; because only few messages are needed as compared to strong consistency guarantees. The performance level can be increased by providing extensive caching mechanisms which in turn lowers the cost.

Cost-Based Adaptive Concurrency Control (C3)

Cost plays an important role in the cloud environment along with the performance [15]. Consistency leads to high cost, whereas weak consistency leads to high operational costs [16]. In C3 approach, a consistency rationing model is used which categorized the data into three: the first category contains data which require ISR, the second category data require SC and the third category data handled with adaptive consistency. At the data level, specific policy will be defined based on that policy consistency level is selected between 1SR and SC at the time of running. Moreover, C3 is implemented on the top of 1SR, SC and SSI concurrency protocols by utilizing the resources provided by the cloud providers. The update anywhere and full replication procedure are the basis for the C3 system model. The updating of all replicas will be carried out in ISR and SSI transactions using 2PC, while SC transactions only commits at the remote local replicas. The C3 model does not introduce any hindrance for the replication strategy. Each and every replica in the system is known to all other replicas. The C3 procedure uses an adaptive layer, which allows the dynamic switching between the different CCPs at runtime. Thus the reduction of operational costs and transaction response time is possible.

VI. Proposed Methodology

In this paper, we study the SSE for string search. In the SSE, the client encrypts the data and stores it on the cloud. It may be noted that client can organize the data in an arbitrary manner and can maintain additional data structures to achieve desired data efficiently. In this process, the initial client side computation is thus as large as the data, but subsequent computations to access data is less for both client and the cloud server.

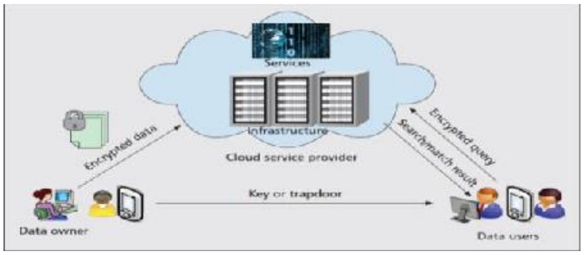

Fig. Proposed Architecture diagram

VII. Conclusion

This research presents a secure search for multiple data owners and multiple data users in the cloud computing environment. Dynamic secret key generation and a new data user authentication algorithms are use to authenticate data users and detect attackers who perform illegal searches. Secure search protocol is use to enable the cloud server to perform secure search among multiple owners data encrypted with different secret keys. We developed a novel method of keyword transformation and introduce the stemming algorithm. With these techniques, the proposed scheme is able to efficiently handle more misspelling mistake. The proposed method introduced ostrovsky scheme to detect data leakage in cloud environment. The proposed schemes enable authenticated data users to achieve secure, convenient, and efficient searches over multiple data owners and detect attackers who steal the secret key and perform illegal hacking. In future work relationships among query keywords have to be considered to enhance the system and introduce the keyword weight to the search protocol design.

References

[1] A New Lightweight Symmetric Searchable Encryption Scheme for String Identification Indranil Ghosh Ray, Yogachandran Rahulamathavan and

Muttukrishnan Rajarajan, Senior Member, IEEE 2168-7161 (c) 2018 IEEE.

[2] Ming Liu and Rafael A. Calvo, “Secure ranked keyword search over encrypted cloud data,” in Proc. IEEE 30th Int. Conf. Distrib. Comput. Syst. (ICDCS), Jun. 2012, pp. 253–262, 2012.

[3] Ning Cao, Cong Wang[2014], “PrivacyPreserving Multi-keyword Ranked Search over Encrypted Cloud Data” in Proc. Of EDBT, 2014, pp. 287–298, 2014.

[4] Q. Wang, C. Wang, K. Ren, W. Lou, and J. Li, “Enabling public auditability and data dynamics for storage security in cloud computing,”IEEE Trans. Parallel Distrib. Syst., vol. 22, no. 5, pp. 847–859, 2009.

[5] Reza Curtmola[2006], “Efficient multi-keyword ranked query over encrypted data in cloud computing,” Future Generat. Comput. Syst., vol. 30, pp. 179–190, 2006.

[6] RuixuanLi Zhiyong Xu [2014] “Efficient multikeyword ranked query over encrypted data in cloud computing” IEEE Trans.Parallel Distrib. Syst., vol. 25, no. 1, pp. 222–233, Jan. 2014.

[7] R.V.V Murali Krishna and Ch. Satyananda Reddy, “Privacy-preserving multi-keyword text search in the cloud supporting similarity-based ranking,” in Proc. 8th ACM SIGSAC Symp. Inf., Comput. Commun. Secur., 2012, pp. 71–82.

[8] P.Sridevi et al [2017] “An Efficient EncryptionThen-Compression System using Asymmetric Numeral Metho”, International Journal of Engineering and Technology (IJET), Vol 9 No 5 OctNov 2017, ISSN (Print) : 2319-8613,ISSN (Online) : 0975-4024.

[9] Z. Xia, X. Wang, X. Sun, and Q. Wang, “A secure and dynamic multikeywordranked search scheme over encrypted cloud data,” IEEE Trans.Parallel Distrib. Syst., vol. 27, no. 2, pp. 340–352, 2015.

[10] Zhihua Xia,“A Secure and Dynamic Multikeyword RankedSearch Scheme over Encrypted Cloud Data”, Ieee Transactions On Parallel And Distributed Systems Vol: Pp No: 99 Year 2015.

Authors

D. Vijaya Durga is pursing M.TECH (CSE) in the Department of Computer Science and Engineering from BVC Institute of Technology & Science, Batlapalem, Amalapuram, AP, India.

A.P.V.D.L. Kumar is working as Associate Professor in Department of Computer Science & Engineering, BVC Institute of Technology & Science, Batlapalem, Amalapuram, AP, India.