



## A Survey on Privacy and integrity preserving Boolean query processing on encrypted cloud data

<sup>1</sup>Surineni Monisha,

<sup>2</sup>K.Malathi,

<sup>1</sup>Student, <sup>2</sup>Assistant Professor,

<sup>1,2</sup>Department of Computer Science and Engineering,

<sup>1,2</sup>Saveetha School of Engineering

<sup>1,2</sup>Saveetha University, Chennai.

<sup>1</sup>monishasurneni1997@gmail.com, <sup>2</sup>malathi@saveetha.com

**Abstract**---Nowadays, Cloud storage is gaining importance among individual and institutional users. Cloud server is always considered to be un-trusted, thus the data stored in cloud is always encrypted and uploaded. Existing schemes available on query processing over cloud data, they explain both Boolean and K-NN methods on cloud. However, those work lag under verifying results. The proposed work propose an efficient scheme for verifying the query results. In the proposed work, an accumulation tree is constructed based on encrypted data and then outsourced both of them to the cloud. Data users, search with a Boolean query, the cloud service provider generates the matching proof according to the result by mapping Boolean query operations to set operations while keeping privacy-preservation and accomplishing the verification requirements.

**Keywords:** Data Privacy, Encrypted Cloud Data, Boolean Query, K-NN Query, Out-sourced Data, Authentication.

### I. INTRODUCTION

Cloud infrastructure and flexibility motivate institutions and individuals to store their records to cloud servers. Data owners store their records to cloud server, these data can be accessed by data owner or authenticated users by querying on cloud data. However, the privacy and confidentiality concerns of data often make them reluctant to do that. To preserve the data privacy and confidentiality, outsourced data is encrypted by data owner before outsourcing to the cloud. Thus the records in cloud storage available will be safe. However, since data in the cloud is encrypted, thus, it is necessary to propose an effective system to obtain encrypted data from the cloud is very imperative for outsourcing storage applications. Existing works such as searchable symmetric encryption (SSE) available for searching encrypted data over cloud. Using these techniques, data owner can outsource encrypted data to safeguard the privacy and confidentiality of data, and data users can query encrypted data the cloud.

### II. PROBLEM STATEMENT

As the data owner may dealt with complex data such as patient records, even some situation, the query from user may be used for early detection of cancer. In such scenario, it is very significant to ensure the verifiability of data retrieved from cloud server. It is must for the cloud server to have the ability to prov-

ide the proof with which the data owner/users are able to validate the reliability of the search result returned by the cloud server. The query result proof verification process should be secure and privacy-preserving.

To realize query integrity verification in DSSE, query operations is mapped to set operations. The query verification can be realized by using the accumulation tree. The problem to be resolved on these criteria would be mapping in DSSE. Constructing the accumulation tree with limited information of DSSE. Achieving the secure and privacy-preserving query integrity verification in DSSE while ensuring the verification requirements.

### III. LITERATURE SURVEY

**S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, Proceedings in 2013.** In the setting of searchable symmetric encryption (SSE), a data owner D outsources a database (or record/document accumulation) to a remote server E in encoded frame to such an extent that D can later pursue the gathering at E while hiding data about the records and queries from E. Leakage to E is to be bound to very much characterized types of information access and query patterns while avoiding divulgence of explicit data and query plaintext values. As of late, Cash et al. exhibited a protocol, OXT, which can run self-assertive boolean inquiries in the SSE setting and which is strikingly productive notwithstanding for

huge databases. In this paper we research a richer setting in which the data owner  $D$  outsources its information to a server  $E$  yet  $D$  is currently intrigued to permit customers (outsiders) to look through the database to such an extent that customers take in the data  $D$  approves them to learn yet nothing else while  $E$  still does not find out about the information or questioned esteems as in the essential SSE setting.

**S.Kamara, C. Papamanthou, and T. Roeder, in 2012.** Searchable symmetric encryption (SSE) permits a customer to encrypt data in this kind of manner. The maximum on the spot utility of SSE is cloud storage, wherein it allows a purchaser to securely outsource information to an untrusted cloud service provider without sacrificing the capability to look over it. SSE provided a large number of schemes that attain numerous levels of safety and efficiency have been proposed. Any realistic SSE scheme, but, ought to (at a minimum) fulfill the subsequent properties: sublinear search time, security in opposition to adaptive chosen-key-word spasms, condensed indexes and the capacity to delete and add files correctly. unluckily, not one of the formerly-known SSE structures accomplish all these properties alike. This critically restricts the practical value of SSE and reduces its risk of deployment in real-international cloud storage structures.

**D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, in 2013.** This work gives the design and evaluation of searchable symmetric encryption (SSE) on outsourced encrypted data and that scales to very big databases and arbitrarily-structured statistics which includes loose textual content seek. thus far, paintings in this vicinity has targeted in particular on single-keyword search. For the case of conjunctive seek, previous SSE structures required work linear inside the overall quantity of documents within the database and furnished exact privateness best for established characteristic-value records, rendering those solutions too gradual and inflexible for big practical databases.

**M. Naveed, M. Prabhakaran, and C. A. Gunter, in 2014.** Dynamic Searchable Symmetric Encryption allows user to store and then immediate keyword searches on those encrypted files, while enlightening negligible statistics to the server. Greater inexperienced than contemporary schemes even as revealing less information to the server than previous schemes, accomplishing fully adaptive safety in opposition to honest-but-curious servers. They carried out a sample of this scheme and verified its performance on datasets from previous work. Aside from its tangible efficiency, scheme is likewise less

difficult: especially, it does not longer required by the server to help any operation aside from add and down load of facts. hence the server in this scheme may be based totally on a cloud carrier, in preference to a cloud computation carrier as properly, as in earlier paintings. In developing this dynamic SSE scheme, a new primitive known as Blind Storage, which allows users to store a sets of data on a cloud server on a far flung server in one of these manner, that the server doesn't need what number of files are stored, or the size of the each files, as each file is regained, the server acquires about its existence (and might word the same document being downloaded subsequently), however the file's name and contents aren't discovered. that is a primitive with numerous programs apart from SSE, and is of unbiased interest.

**E. Stefanov, C. Papamanthou, and E. Shi, IACR, in 2013.** Dynamic Searchable Symmetric Encryption (DSSE) permits a customer to encrypt his file collection in a manner that it's far still searchable and correctly updatable. But, all DSSE buildings which have been accessible in the literature so far include several problems: both they leak a huge amount of information (e.g., hashes of the key phrases contained within the up to date file) or are inefficient in phrases of space or search/update time (e.g., linear inside the number of files). On this paper we revisit the DSSE hassle. We endorse the primary DSSE scheme that achieves the first-class of both worlds, i.e., each small leakage and efficiency. Specially, our DSSE scheme leaks notably less facts than any other preceding DSSE creation and helps each updates and searches in sublinear time within the worst case, keeping on the identical time a records shape of handiest linear length. We sooner or later offer an implementation of our construction, showing its realistic efficiency.

**S. Nath and R. Venkatesan, in 2013.** Outsourcing data streams and preferred computations to a third celebration inclusive of the cloud is a appropriate choice to many agencies. however, statistics outsourcing and far flung computations intrinsically increase problems of accept as true with, making it important to verify results back by using 0.33 parties. in this context, we advocate a singular solution to verify outsourced grouped aggregation queries (e.g., histogram or square organization-by way of queries) that are commonplace in lots of enterprise programs. We recollect a setting in which a records proprietor employs an untrusted faraway server to run continuous grouped aggregation queries on a information move it forwards to the server. Untrusted clients then question the server for results and successfully verify correctness of the results via the

use of a small and clean-to-compute signature provided through the records proprietor. Our work complements previous works on authenticating far flung computation of choice and aggregation queries. The maximum critical thing of our answer is that it's far publicly verifiable - unlike maximum previous works, we support untrusted clients (who can collude with other customers or with the server). Experimental results on real and synthetic information show that our answer is practical and efficient.

**H. Pang, j. Zhang, and k. Mouratidis, in 2009.** Query solutions from servers activated by by third parties need to be proved, as the third parties may not be depended on or their servers may be negotiated. maximum of the prevailing authentication strategies assemble validity verifications. The MHT, but, imposes intense concurrency constraints that gradual down data updates. They introduced a protocol, constructed upon signature. The protocol offers the critical belongings of permitting new data to be disseminated without delay, even as ensuring that previous values elsewhere a pre-set age can be detected. An efficient confirmation method for ad-hoc equi joins. Similarly, for cloud servers that need to procedure query workloads, a mechanism that extensively reduces the evidence production time by way of caching only a small quantity of strategically selected mixture signatures. The performance and efficacy of our proposed mechanisms are confirmed via giant experiments.

**F. Li, M. Hadjieleftheriou, k. Kollios, and L. Reyzin, in 2006.** Database owner publishes its information via some of faraway servers, with the aim of enabling customers at the threshold of the community to get admission to and query the information extra efficaciously. As servers is probably untrusted or may be compromised, question authentication turns into an crucial factor of ODB systems. Present solutions for this trouble listen totally on static scenarios and are primarily based on idealistic properties for sure cryptographic primitives. on this paintings, first we outline a spread of important and realistic fee metrics related to ODB structures. Then, we logically assess a number of one-of-a-kind procedures, in look for a solution that best influences all metrics. most importantly, look at solutions that can handle dynamic situations, where proprietors periodically replace the statistics dwelling on the servers. sooner or later, we talk question freshness, a new measurement in facts authentication that has not been explored earlier than. Our findings show off the enhanced overall performance appreciably over present procedures.

**F. Li, K. Yi, M. Hadjieleftheriou, and G. Kollios, in 2007.** As computer structures are vital additives of many critical business offerings, the want for cozy online transactions is now turning into glaring. The claim for such applications, because the market place grows, exceeds the capability of individual companies to offer fast and dependable services, making outsourcing skills a key participant in alleviating problems of scale. don't forget a inventory broking that needs to provide a real-time stock buying and selling monitoring carrier to clients. since the fee of multicasting this information to a large audience would possibly turn out to be prohibitive, the broking should outsource the inventory feed to 0.33-party vendors, who are in flip re- sponsible for forwarding the perfect sub-feed to customers. obviously, in vital packages the integrity of the 1/3- celebration ought to not be taken without any consideration. on this work we observe a diffusion of authentication algorithms for choice and aggregation queries over sliding home windows. Our algorithms en- in a position the end-customers to prove that the outcomes provided by means of the third-birthday party are accurate, i.e., equal to the consequences that might were computed by means of the authentic issuer. Our solutions are based on Merkle hash trees over a forest of area par- titioning information systems, and attempt to leverage key features, like replace, question, signing, and authentication costs. We present targeted theoretical evaluation for our answers and empirically examine the proposed strategies.

**S. Papadopoulos, Y. Yang, and D. Papadias, in 2007.** We have a look at processing and authentication of long-walking queries. On this scenario, a statistics owner (DO) constantly transmits its information to a service company (SP), collectively with additional authentication facts. customers check in continuous range queries to the SP. each time the statistics trade, the SP ought to replace the effects of all affected queries and inform the clients hence. The customers can affirm the correctness of the results using the authentication information provided by means of the DO. as compared to conventional databases, move environments pose new challenges together with the want for instant structure updating, assist for continuous processing and authentication. Particularly, similarly to the correctness of character results, the client need to be able to verify that there aren't any missing consequences in between updates.

**K. Yi, F. Li, G. Cormode, M. Hadjieleftheriou, G. Kollios, and D. Srivastava, in 2009.** Due to the overpowering flow of information outsourcing, is a

natural and powerful paradigm for businesses to cope with the difficulty of scale. Within the standard data outsourcing model, the statistics proprietor outsources streaming records to 1 or greater 1/3-birthday celebration servers, which solution queries posed by way of a doubtlessly huge number of customers on the records proprietor's behalf. statistics outsourcing intrinsically raises troubles of believe, making outsourced question assurance on facts streams a hassle with critical realistic implications. present solutions including signatures and collision-resistant hash features, which provides simplest work for sure varieties of queries, as an instance, simple selection/aggregation queries.

**Q. Zheng, S. Xu, and G. Ateniese, in 2014.** It's miles common these days for data owners to store data to the cloud. This however brings a number problems, including: How have to a records owner supply seek competencies to the information customers? How can the legal records customers seek over a information owner's outsourced encrypted information? How can the information users be confident that the cloud faithfully completed the hunt operations on their behalf? motivated by way of these questions, we suggest a novel cryptographic answer. The answer permits a information user, whose credentials satisfy a information owner's get admission to manipulate coverage, to (i) search over the records proprietor's outsourced encrypted data, (ii) outsource the tedious seek operations to the cloud, and (iii) confirm whether the cloud has faithfully accomplished the search operations.

**C. Papamanthou, R. Tamassia, and N. Triandopoulos, in 2008.** Hash tables are essential statistics structures that optimally solution club

queries. Assume a purchaser shops  $n$  factors in a hash desk this is outsourced at a remote server so that the consumer can keep area or attain load balancing. Authenticating the hash table capability, i.e., authenticating the accuracy of queries replied by the cloud server and making sure the integrity, lying outdoor the administrative control of the patron, may be malicious.

**R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, in 2006.** Searchable symmetric encryption (SSE) licenses a party to outsource its statistics to some other birthday party (a server) in a non-public way, as maintaining the capability to selectively seek over it. This hassle has been the point of interest of active research in latest years. in this paper we show solutions to SSE that concurrently revel in the subsequent properties: each solutions are greater efficient than all preceding steady-spherical schemes. specifically, the work accomplished by using the server per lower back report is consistent as opposed to linear in the size of the records.

**D. Boneh and M. Franklin, in 2001.** It proposes a fully determined identification-based encryption scheme (IBE). The scheme has selected ciphertext protection inside the accidental vision model assuming an elliptic curve variant of the computational Diffie-Hellman problem. Our product is based on the Weil pairing. We supply specific definitions for at ease identification based encryption schemes and supply numerous applications for such structures. Finally, I conclude the work of literature survey with their methodologies comparison as shown in the Table 1 (Methodology Comparison).

**Table 1: Methodology Comparison**

| S.No | Author/Year                                                                   | Methodology                                                      | Merits                                                            | Demerits                                                                                                                                      |
|------|-------------------------------------------------------------------------------|------------------------------------------------------------------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|
| 1    | S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, Proceedings, 2013 | Searchable symmetric encryption (SSE) technique is used          | Hiding information about the database and queries from attackers. | Sublinear search time, security against adaptive chosen-keyword attacks, compact indexes and the ability to add and delete files efficiently. |
| 2    | S. Kamara, C. Papamanthou, and T. Roeder, Proceedings, 2012                   | Dynamic searchable symmetric encryption (DSSE) technique is used | Encrypt its data in a way that this data can still be searched.   | Single-keyword search only enabled                                                                                                            |

|   |                                                                           |                                                                                                                                                           |                                                            |                                                                                                                                                                                                            |
|---|---------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 3 | D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, 2013 | Design and analysis of searchable symmetric encryption (SSE) protocol is changed to general Boolean queries on outsourced symmetrically encrypted data    | Good privacy                                               | Structured attribute-value solves too slow                                                                                                                                                                 |
| 4 | M. Naveed, M. Prabhakaran, and C. A. Gunter, 2014.                        | Dynamic SSE scheme is more efficient than existing schemes while revealing less information to the server                                                 | Keyword searches reveals minimal information to the server | Data leakage is high                                                                                                                                                                                       |
| 5 | E. Stefanov, C. Papamanthou, and E. Shi, IACR, 2013.                      | Dynamic Searchable Symmetric Encryption (DSSE) proposed to encrypt his document collection in a way that it is still searchable and efficiently updatable | Small leakage and efficiency                               | Due to the nature of the delegation/outsourcing, the cloud server can fully control the outsourced data and decide the SSE query result of hospitals or research institutes, which causes issues of trust. |

#### IV. SYSTEM ARCHITECTURE

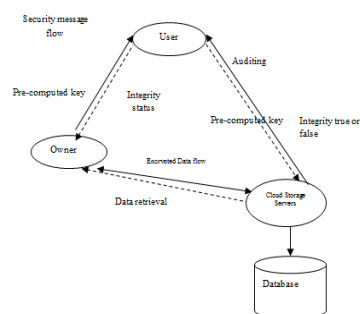


Fig.1: System Architecture

#### V. PROPOSED SYSTEM

To present a publicly query result verifiable scheme, which can publicly verify whether the cloud has faithfully executed Boolean search operations in dynamic searchable symmetric encryption (DSSE). To propose a publicly Boolean query integrity verification scheme over the outsourced dynamic encrypted data to check the authenticity is represented in Fig.1, freshness and completeness of

the query result. To construct a special Merkle hash tree named accumulation tree in DSSE to map Boolean operations of keywords to set operations, which is different from the traditional signature or aggregate signature.

##### a. PROPOSED ALGORITHM

Input: Data document 'D', User query 'Q'

Output: Encrypted content 'E' and proof of verification 'T' or 'F'

Step 1: Data owner encrypts outsourced documents and generates the document index for SSE

Step 2: Data owner picks a constant  $\epsilon$ , where  $0 < \epsilon < 1$ .

Step 3: Data owner constructs tree AT according to  $\text{stag}(w_i)$  that has  $l = \lceil 1/\epsilon \rceil$  levels and  $m$  leaves, where  $m$  is the number of  $W$ .

Step 4: DO sets  $d_0 = d(r)$  where  $r$  is the root of AT and keeps it.

Step 5: Finally, DO outsources AT with encrypted DB to the cloud.

Step 6: User sends input key  $K$  and proof key  $K_c$  for a data  $D$ , signature  $S$  for checking the correctness of data block specified.

Step 7: Server takes input key  $K$  and  $K_c$  for a data  $D$ , signature  $S$ . It results a data integrity proof  $P$  for the block of data specified

Step 8: Output is TRUE if data correctness exists FALSE otherwise

## VI. CONCLUSION

The problem of verifying the authenticity, freshness, and inclusiveness of the Boolean query result over the outsourced encrypted data. A publicly query result verifiable scheme is proposed by constructing the accumulation tree to achieve the query integrity verification while keeping privacy-preserving and efficiently practical. The analysis shows that without protecting the access pattern, our scheme can keep the privacy-preserving of private information retrieval. The performance demonstrates our scheme is scalable.

## REFERENCES

- [1] S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, "Outsourced symmetric private information retrieval," in Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security. ACM, 2013, pp. 875–888.
- [2] S. Kamara, C. Papamanthou, and T. Roeder, "Dynamic searchable symmetric encryption," in Proceedings of the 2012 ACM conference on Computer and communications security. ACM, 2012, pp. 965–976.
- [3] D. Cash, S. Jarecki, C. Jutla, H. Krawczyk, M. Rosu, and M. Steiner, "Highly-scalable searchable symmetric encryption with support for boolean queries," in Advances in Cryptology-CRYPTO 2013. Springer, 2013, pp. 353–373.
- [4] M. Naveed, M. Prabhakaran, and C. A. Gunter, "Dynamic searchable encryption via blind storage," in Proceedings of the IEEE Symposium on Security and Privacy, San Jose, CA, May, 2014.
- [5] E. Stefanov, C. Papamanthou, and E. Shi, "Practical dynamic searchable encryption with small leakage," IACR Cryptology ePrint Archive, vol. 2013, p. 832, 2013.
- [6] <http://www.wired.com/2009/01/magnolia-suffer/>.
- [7] <http://mashable.com/2011/02/27/gmail-glitch/>.
- [8] S. Nath and R. Venkatesan, "Publicly verifiable grouped aggregation queries on outsourced data streams," in Proceedings of the 29th International Conference on Data Engineering (ICDE). IEEE, 2013, pp. 517–528.
- [9] H. Pang, J. Zhang, and K. Mouratidis, "Scalable verification for outsourced dynamic databases," Proceedings of the VLDB Endowment, vol. 2, no. 1, pp. 802–813, 2009.
- [10] F. Li, M. Hadjieleftheriou, K. Kollios, and L. Reyzin, "Dynamic authenticated index structures for outsourced databases," in Proceedings of the 2006 ACM SIGMOD international conference on Management of data. ACM, 2006, pp. 121–132.
- [11] F. Li, K. Yi, M. Hadjieleftheriou, and G. Kollios, "Proof-infused streams: Enabling authentication of sliding window queries on streams," in Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, 2007, pp. 147–158.
- [12] S. Papadopoulos, Y. Yang, and D. Papadias, "Cads: Continuous authentication on data streams," in Proceedings of the 33rd international conference on Very large data bases. VLDB Endowment, 2007, pp. 135–146.
- [13] K. Yi, F. Li, G. Cormode, M. Hadjieleftheriou, G. Kollios, and D. Srivastava, "Small synopses for group-by query verification on outsourced data streams," ACM Transactions on Database Systems (TODS), vol. 34, no. 3, p. 15, 2009.
- [14] Q. Zheng, S. Xu, and G. Ateniese, "Vabks: Verifiable attribute-based keyword search over outsourced encrypted data," in Proceedings of the 2014 INFOCOM 2014. IEEE, 2014.
- [15] C. Papamanthou, R. Tamassia, and N. Triandopoulos, "Authenticated hash tables," in Proceedings of the 15th ACM conference on Computer and communications security. ACM, 2008, pp. 437–448.



