Dual Server Multi Keyword Search Over Encrypted Spatial Data in Cloud Storage

Dr.S.Sadesh[1], G.Ashish[2]
[1]Professor, [2]M.E Student, Dept. Computer Science and Engineering,
Velalar College of Engineering and Technology, Erode 12, Tamil Nadu, India.

***Abstract*—The advent of cloud computing, data owners are motivated to outsource their complex data management systems from local sites to commercial public cloud for great flexibility and economic savings. But for protecting data privacy, sensitive data has to be encrypted before outsourcing, which obsoletes traditional data utilization based on plaintext keyword search. Thus, enabling an encrypted cloud data search service is of paramount importance. Considering the large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. Related works on searchable encryption focus on single keyword search or Boolean keyword search, and rarely differentiate the search results. In this paper, for the first time, we define and solve the challenging problem of privacy-preserving multi-keyword ranked ontology keyword mapping and search over encrypted cloud data (DSMK), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "Dual Server Multi keyword data search", i.e., as many matches as possible, to capture the similarity between search query and data documents, and further use "inner product similarity" to quantitatively formalize such principle for similarity measurement. We first propose a basic DSMK scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset further show proposed schemes indeed introduce low overhead on computation and communication.**

***Keywords*— Fine-grained inquiry, Spatial Data, Tokens, , Sub tree.**

I. INTRODUCTION

The Generic Privacy Preserving Problem

The problem of learning something without revealing ones own data is not new. It was proposed way back in 1982 by Yao [5]. It has become very important as data has started to grow a million times faster and along with it the demands to keep it for oneself only. When the problem was proposed the web had just come out of its infancy, today it is mature and large and spread to the remotest corners of the world. The authors marked back then that explosive progress in networking, storage, and processor technologies has led to the creation of ultra large databases that record unprecedented amount of transactional information [1]. We are more concerned in privacy

Grenze ID: 02. ICIEICE.2021.1.25

preservation with context to data mining algorithms. This is one point where the privacy can be trapped. Suggestion from paper that data mining and data warehousing go hand-in- hand: Most tools operate by gathering all data into a central site, then running an algorithm against that data [6]. However, privacy concerns can prevent building a centralized warehouse— data may be distributed among several custodians, none of which are allowed to transfer their data to another site. It should be noted that what data mining algorithms produce is knowledge, and that data mining results rarely violate privacy, as they generally reveal high-level knowledge rather than disclosing instances of data. However, the concern among privacy advocates is well founded, as bringing data together to support data mining makes misuse easier. The problem is not data mining, but the way data mining is done [7]. 1. The Solutions to the problem PPDM is a new era of research in data mining, where data mining algorithms are analysed for possible infringement in privacy. PPDM research usually takes one of the three philosophical approaches: (1) data hiding, in which sensitive raw data like identifiers, name, addresses, etc. were transformed, jammed, or trimmed out from the original database, in order for the users of the data not to be able to compromise another person's privacy; (2) secure multiparty computation, where distributed data are encrypted before released or shared for computations; and (3) rule hiding, in which sensitive knowledge extracted from the data mining process be excluded for use, because private information may be derived from the released knowledge; thus, no party knows no matter which except its own inputs and the results. The crucial goal of PPDM is to develop efficient algorithms that allow one to extract relevant knowledge from a large amount of data, while prevent sensitive data and information from leak or deduction [10]. The third approach can be broadly called cryptographic approach to solve PPDM problems. A very nicely classifies all the proposed solutions into various categories depending on what methods are used in [8]. Various ways to handle PPDM problems including the cryptographic approaches available in [4][3]. 2. The Quantifiers of efficient solution The most general parameters for analysing efficient solution are overall performance in all the areas. A framework is but required to get exact comparative measures. Attempts have been made in past to generalize a framework. One is proposed by Bertino et al. The framework they identified was based on the following evaluation dimensions [9]: -- Efficiency. The ability of a privacy preserving algorithm to execute with good performance in terms of all the resources implied by the algorithm; – Scalability. This factor evaluates the efficiency trend of a PPDM algorithm for increasing sizes of the data from which relevant information is mined while ensuring privacy; – Data quality after the application of a privacy preserving technique. Considered both as the quality of data themselves and the quality of the data mining results after the hiding strategy is applied; – Hiding failure. The portion of sensitive information that is not hidden by the application of a privacy preservation technique; – Privacy level offered by a privacy preserving technique. It estimates the degree of uncertainty, according to which sensitive information can still be predicted even if it has been hidden. Such framework allows one to assess the different features of a privacy preserving algorithm according to a variety of evaluation criteria. 3 Privacy Preserving Techniques the most used technique yet has been secure multiparty computation. As mentioned earlier, the basic privacy preservation problem is a classical multiparty problem. Cryptography-based SMC has the highest accuracy in data mining and good privacy preservation capability as well; however, it has strict usages as it is only applicable to a distributed data environment [6].

Cloud Based Privacy Data Sharing Using Datamining:

The rapid expansion of data, the data owners tend to store their data into the cloud to release the burden of data storage and maintenance [1]. However, as the cloud customers and the cloud server are not in the same trusted domain, our outsourced data may be under the exposure to the risk. Thus, before sent to the cloud, the sensitive data needs to be encrypted to protect for data privacy and combat unsolicited accesses. Unfortunately, the traditional plaintext search methods cannot be directly applied to the encrypted cloud data any more. The traditional information retrieval (IR) has already provided multi-keyword ranked ontology keyword mapping and search for the data user. In the same way, the cloud server needs provide the data user with the similar function, while protecting data and search privacy.

II. Literature Survey

Facilitating secure and efficient spatial query processing on the cloud

Database redistributing is a typical distributed computing worldview that permits information proprietors to exploit its on-request capacity and computational assets. Existing methodologies either bargain classification of the information or experience the ill effects of high correspondence cost between the worker and the client. The client issues scrambled spatial range questions to the specialist organization and afterward utilize the encryption key to

unscramble the inquiry reaction returned [11-18]. This permits a harmony between the security of information and productive inquiry reaction as the inquiries are handled on encoded information at the cloud worker. In addition, we contrast and existing methodologies on enormous datasets and show that this methodology lessens the normal question correspondence cost between the approved client and specialist co-op, as just a solitary round of correspondence is required by the proposed approach. We characterize a few assault models and show that our plan gives solid protection from them. This permits a harmony between the security of information and quick reaction time as the inquiries are handled on encoded information at the cloud worker. Along these lines, the double change strategy protects the information as well as empowers the validated clients to recover spatial range question reactions effectively.

Fastgeo: efficient geometric range queries on encrypted spatial data

The rising interest of re-appropriating information is moving huge scope datasets, including enormous scope spatial datasets, to open mists. In this paper, we formalize the idea of Geometrically Searchable Encryption, and propose an effective plan, named FastGeo, to ensure the security of customers' spatial datasets put away and questioned at an open worker. FastGeo underpins subjective mathematical zones, accomplishes sub linear search time, and empowers dynamic refreshes over scrambled spatial datasets. We propose FastGeo, an effective two-level hunt plot that can work mathematical ranges over encoded spatial datasets. Our test results over a real-world dataset exhibit its viability by and by. Additionally, our correlation with past arrangements shows that the overall thought of two-level inquiry can be utilized as a significant technique to help search time and empower profoundly productive updates over scrambled information when complex activities, for example, process then compare activities, are engaged with search.

Enabling efficient and geometric range query with access control over encrypted spatial data

A basic query function, range query has been exploited in many scenarios such as Sql retrieves, location-based services, and computational geometry. A long-standing problem is that the user's data may be completely revealed to the cloud server because it has full data access right. We propose an Efficient and Geometric Range Query scheme (EGRQ) supporting searching and data access control over encrypted spatial data. We employ secure KNN computation, polynomial fitting technique and order-preserving encryption to achieve secure, efficient and accurate geometric range query over cloud data. To improve the efficiency, R-tree is adopted to reduce the searching space and matching times in whole search process. Finally, we theoretically prove the security of our proposed scheme in terms of confidentiality of spatial data, privacy protection of index and trapdoor, and the unlink ability of trapdoors. EGRQ can achieve both arbitrary geometric range query and data access control with one round of communication. Benefited from R-tree adopted to reduce the searching space and matching times in whole querying process, it is also not vulgar that the performance of computing overhead in our EGRQ. Security analysis also demonstrates the high security of our proposed scheme in terms of confidentiality of spatial data, privacy protection ofindex and trapdoor, and the unlink ability of trapdoor.

Light weight and privacy-preserving delegable proofs of storage with data dynamics in cloud storage

Distributed storage has been in across the board use these days, which lightens clients' weight of neighbour hood information. Stockpiling. Capacity (POS) is the primary procedure acquainted with address this issue. POS permitting a third gathering to confirm the information trustworthiness in the interest of the information proprietor altogether improves the versatility of cloud administration. we propose another variation plan called "Delegatable Proofs of Storage (DPOS)". Accelerate the label age process by at any rate a few multiple times, without giving up productivity in some other angle. Our plan is sound and protection saving against evaluator in the standard model. The proposed plot is as productive as private key POS plot, particularly very productive in validation label age. Certain POS plots, our own improves the verification label age speed by multiple times. Our conspire likewise forestalls information spillage to the reviewer during the examining procedure.

Secure range search over encrypted uncertainiot outsourced data

The operation of IOT needs a strong data handling capacity, where most of the data is sensor data. Limitations associated with measurement, delays in data updating, and or the need to preserve the privacy of data can result in the sensor data being uncertain. Searchable encryption (SE) scheme is a promising technique that allows the searching over encrypted (uncertain) data stored offshore. Use homo morphic and order-preservin gencryption (OPE) to encrypt data published by the data owners. Design is to ensure the privacy of the dataset, without affecting the efficiency of keyword search on the (encrypted) dataset. The diversity and range of IoT devices will grow as they are deployed in a broader range of applications, ranging from civilian to military and battle field and so on.

The security of uncertain IoT data, particularly those outsourced to the cloud or the edge, we developed an effective indexing technique to support range searches on multidimensional encrypted data. Using the KD-tree to organize the objects to improve the retrieval efficiency. OPE and homomorphic encryption scheme to encrypt the dataset.

Identity-based data outsourcing with comprehensive auditing in clouds

Cloud storage system provides facilitative file storage and sharing services for distributed clients. Identity-based data outsourcing (IBDO) scheme equipped with desirable features advantageous over existing proposals in securing outsourced data. The proxies are identified and authorized with their recognizable identities, which eliminates complicated certificate management in usual secure distributed computing systems. Security analysis and experimental evaluation indicate that our IBDO scheme provides strong security with desirable efficiency. Introduced the notion of identity based data outsourcing and proposed a secure IBDO scheme. The identity-based feature and the comprehensive auditing feature make our scheme advantageous over existing PDP/PoR schemes. Security analyses and experimental results show that the proposed scheme is secure and has comparable performance as the SW scheme.

Verifiable social data outsourcing

Social data outsourcing is an emerging paradigm for effective and efficient access to the social data. A third-party Social Data Provider (SDP) purchases complete social datasets from Online Social Network (OSN) operators and then resells them to data consumers who can be any individuals or entities desiring the complete social data satisfying some criteria. Initiate the study on verifiable social data outsourcing whereby a data consumer can verify the trustworthiness of the social data returned by the SDP. The OSN provider to generate some cryptographic auxiliary information, based on which the SDP can construct a verification object for the data consumer to verify the query-result trustworthiness. Extensive experiments based on a real Twitter dataset confirm the high efficacy and efficiency of our schemes. In this paper, we initiated the study of verifiable social data outsourcing to allow a data consumer to verify the trustworthiness of the social data returned by the SDP. The data consumer to verify the social-graph correctness, social-graph completeness, and content authenticity of any query result returned by an untrusted SDP.

System analysis

Existing system :-The large number of data users and documents in cloud, it is crucial for the search service to allow multi-keyword query and provide result similarity ranking to meet the effective data retrieval need. The searchable encryption focuses on single keyword search or Boolean keyword search, and rarely differentiates the search results.

Disadvantages

- Single-keyword search without ranking is not possible
- Identity based keyword extraction is not available
- Less security.
- Poor reliability.
- Boolean- keyword search without ranking
- Single-keyword search with ranking

Proposed method

We define and solve the challenging problem of privacy-preserving multi-keyword ranked ontology keyword mapping and search over encrypted cloud data (MROS), and establish a set of strict privacy requirements for such a secure cloud data utilization system to become a reality. Among various multi-keyword semantics, we choose the efficient principle of "coordinate matching". We propose the problem of Secured Multi keyword search (SMS) over encrypted cloud data (ECD), Dual server Multi keyword data search and construct a group of privacy policies for such a secure cloud data utilization system. From number of multi-keyword semantics, we select the highly efficient rule of coordinate matching, i.e., as many matches as possible, to identify the similarity between search query and data , and for further matching we use inner data correspondence to quantitatively formalize such principle for similarity measurement.

We first propose a basic Secured multi keyword ranked ontology keyword mapping and search scheme using secure inner product computation, and then improve it to meet different privacy requirements. The Ranked result provides top k retrieval results. Also we propose an alert system which will generate alerts when un-authorized user tries to access the data from cloud, the alert will generate in the form of mail and message.

Advantages of Proposed System

1. Multi-keyword ranked ontology keyword mapping and search over encrypted cloud data **Dual server Multi keyword data search.**
2. "Dual server Multi keyword data search" by inner product similarity.
3. Secured Multi keyword ranked ontology keyword mapping and search: To design search schemes which allow multi-keyword query and provide result similarity ranking for valuable data retrieval, instead of returning undifferentiated results.
4. Privacy: To prevent cloud server from learning additional information from dataset and index, and to meet privacy requirements. Effectiveness with high performance: Above goals on functionality and privacy should be achieved with low communication and computation overhead.

Module description

We define and solve the challenging problem of privacy-preserving multi-keyword ranked ontology keyword mapping and search over encrypted cloud data **(MROS),** and establish a set of **strict privacy** requirements for such a **secure cloud data** utilization system to become a reality.

- Among various multi-keyword semantics, we choose the efficient principle of "**Dual server**". We propose the problem of Secured Multi keyword search (SMS) over encrypted cloud data (ECD), **Dual Server multi keyword data search** and construct a group of privacy policies for such a secure cloud data utilization system.
- From number of **multi-keyword semantics**, we select the highly efficient rule of **coordinate matching**, i.e., as many matches as possible, to identify the similarity between search query and data , and for further matching we use inner **data correspondence** to quantitatively formalize such principle for similarity measurement.
- We first propose a basic Secured multi keyword ranked **ontology keyword mapping** and **search scheme** using secure inner product computation, and then improve it to meet different **privacy** requirements.
- The **Ranked** result provides top k retrieval results. Also we propose an alert system which will **generate alerts** when un-authorized user tries to access the **data from cloud**, the alert will generate in the form of **mail** and **message**.

Encrypt module:-This module is used to help the server to encrypt the document using TRIPLE DES Algorithm and to convert the encrypted document to the Zip file with activation code and then activation code send to the user for download.

Client module:-The user is going to select the required file and register the user details and get activation code in mail from the "customerservice404" email before enter the activation code. After user can download the Zip file and extract that file.

Admin module:-This module is used to help the server to view details and upload files with the security. Admin uses the log key to the login time. Before the admin logout, change the log key. The admin can change the password after the login and view the user downloading details and the counting of file request details on flowchart. The admin can upload the file after the conversion of the Zip file format.

File upload Module:-Admin uses the log key to the login time. Before the admin logout, change the log key. The admin can change the password after the login and view the user downloading details and the counting of file request details on flowchart.

Ontology keyword mapping:-Cloud data under the aforesaid model, our system design should instantaneously achieve security and performance by **(MROS).**

Privacy-preserving:-To prevent the cloud server from learning additional information from the dataset and the index, and to meet privacy.

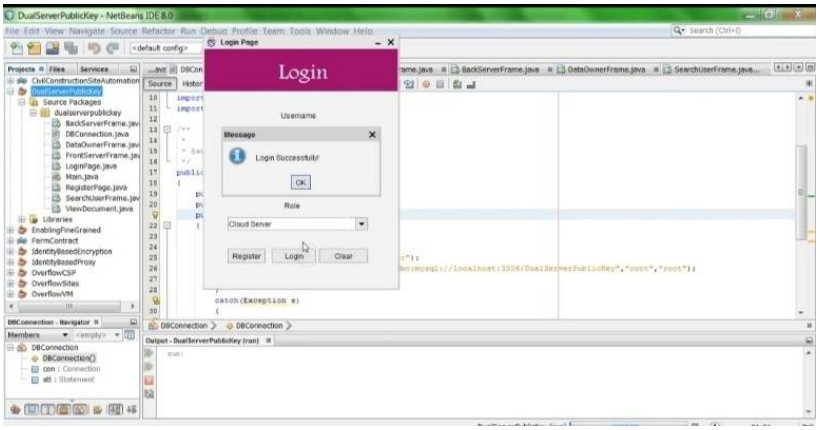
Efficiency:-Above goals on functionality and privacy should be achieved with low communication and computation over head.

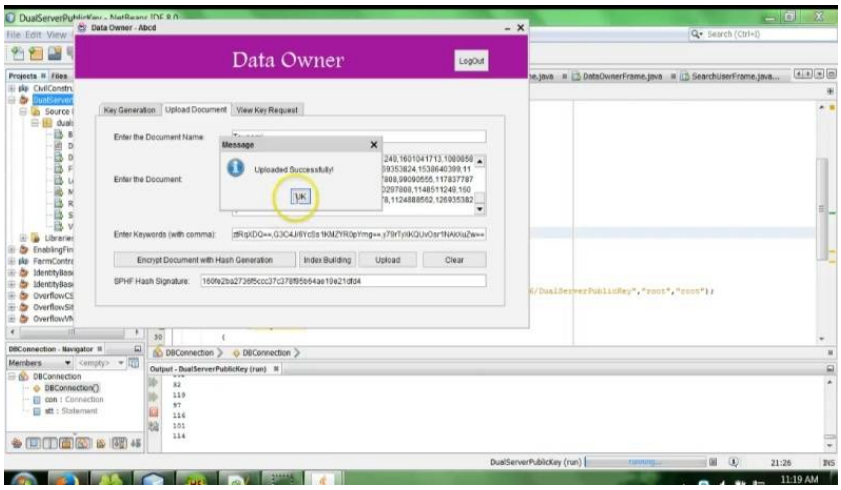
Coordinate matching:-"Dual Server Multi keyword data search" is an intermediate similarity measure which uses the number of query keywords appearing in the document to quantify the relevance of that document to the query. When users identify the exact subset of the dataset to be regained, Boolean queries achieve well with the exact search necessity stated by the user. It is more elastic for users to identify a list of keywords indicating their concern and regain the most relevant documents with a rank order.

- Data privacy
- Index privacy
- Keyword Privacy.
- The trapdoor can be generated in a cryptographic way to protect the query keywords

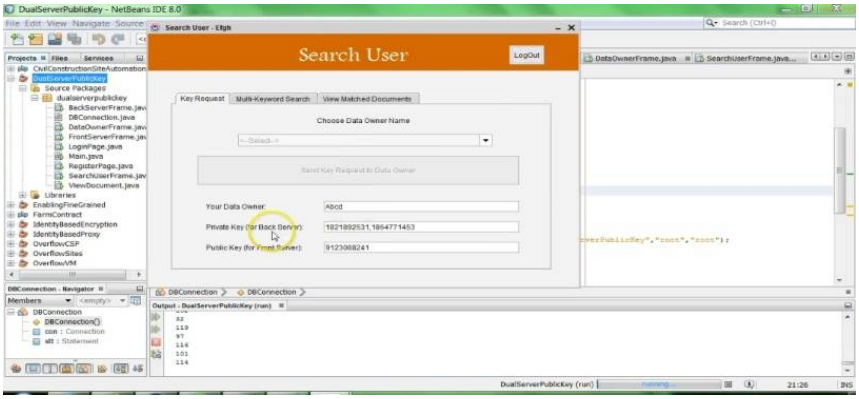
III. Result And Implementation

Login page

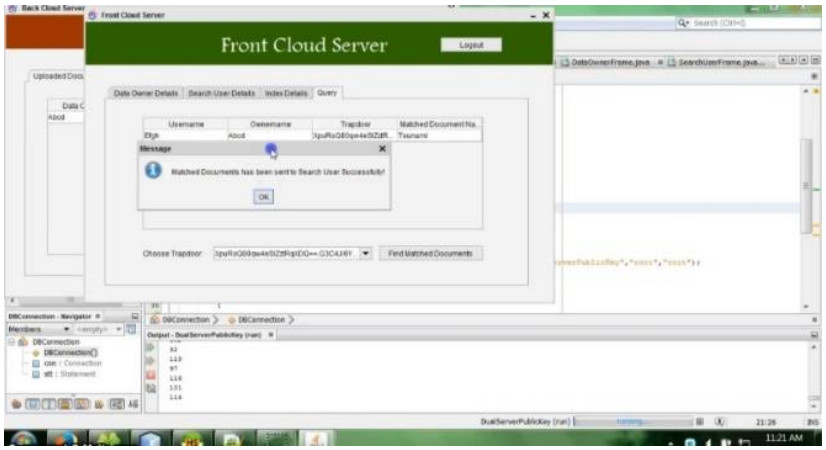


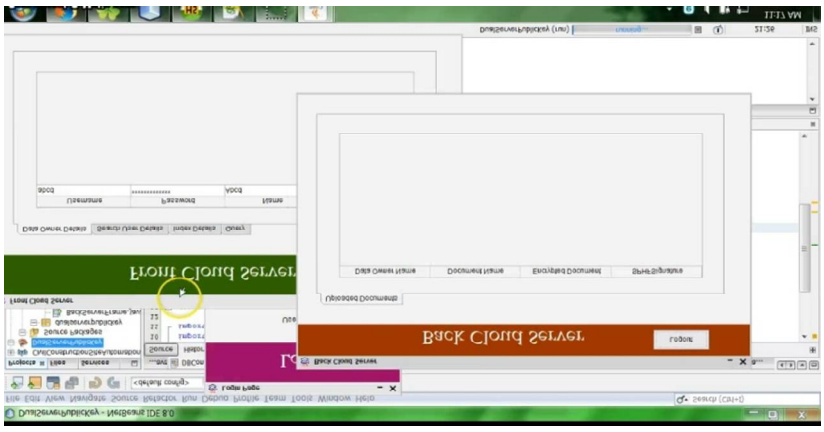


Search user

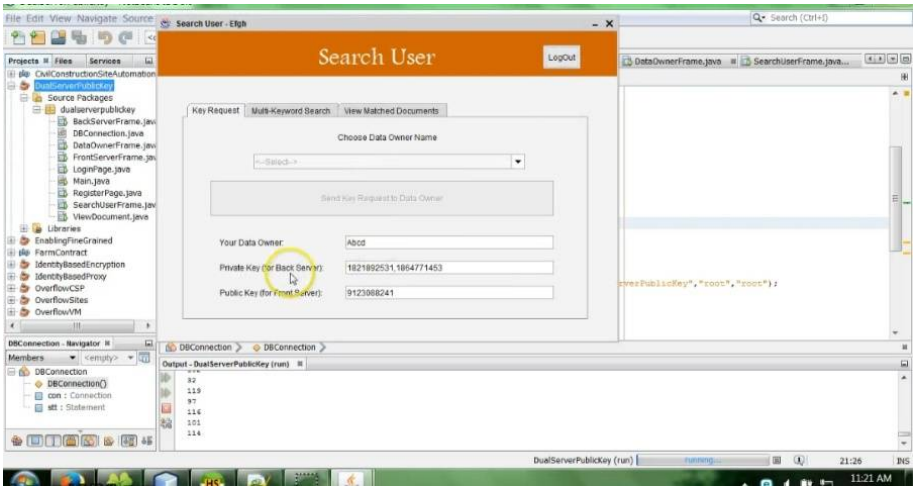


Cloud server

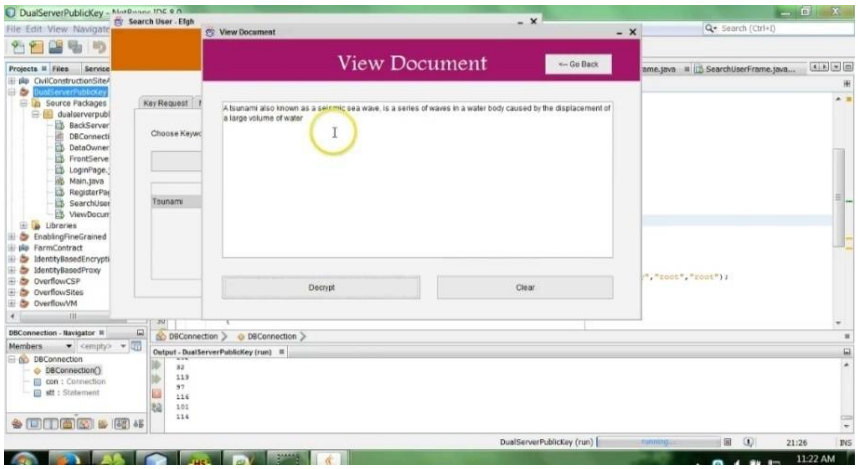


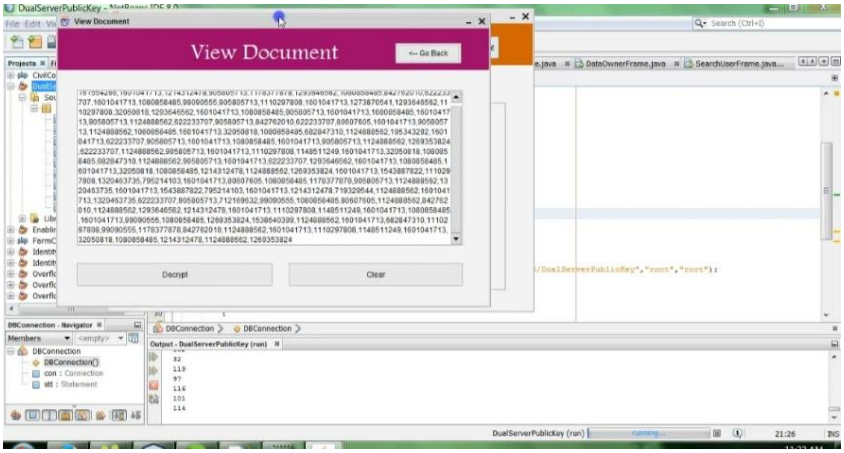


Search page



View document page





IV. CONCLUSION

In this paper, for the first time we define and solve the problem of multi-keyword ranked search over encrypted cloud data, and establish a variety of privacy requirements. Among various multi-keyword semantics, we choose the efficient principle of "Dual Server Multi keyword data search", i.e., as many matches as possible, to effectively capture similarity between query keywords and outsourced documents, and use "inner product similarity" to quantitatively formalize such a principle for similarity measurement For meeting the challenge of supporting multi-keyword semantic without privacy breaches, we first propose a basic MRSE scheme using secure inner product computation, and significantly improve it to achieve privacy requirements in two levels of threat models. Thorough analysis investigating privacy and efficiency guarantees of proposed schemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead on both computation and communication. As our future work, we will explore supporting other multi-keyword semantics (e.g., weighted query) over encrypted data, integrity check of rank order in search result and privacy guarantees in more stronger threat model.

REFERENCES

[1] Feng Tian, Zhenqiang Wu, Xiaolin Gui,Jianbing Ni, "Fine grained query authorization with integrity veritification over encrypted spatial data in cloud storage", IEEE Transctions on Cloud Computing, 2020,DOI:10.1109/TCC.2020.3010915

[2] M. Talha, I. Kamel, and Z. A. Aghbari, "Facilitating secure and efficient spatial query processing on the cloud," IEEE Transactions on Cloud Computing, 7(4), 2019.

[3] Y. Miao, J. Ma, X. Liu, J. Weng, H. Li, and H. Li, "Lightweight fine-grained search over encrypted data in fog computing," IEEE Transactions on Services Computing, 12(5), 2019, 772–785,.
[4] Wang, M. Li, and L. Xiong, "Fastgeo: Efficient geometric range queries on encrypted spatial data," IEEE Transactions on Dependable and Secure Computing, 16(2), 2019, 245–258.
[5] G. Xu, H. Li, Y. Dai, K. Yang, and X. Lin, "Enabling efficient and geometric range query with access control over encrypted spatial data," IEEE Transactions on Information Forensics and Security, 14(4), 2019, 870–885.
[6] Y. Ji, C. Xu, J. Xu, and H. Hu, "vabs: Towards verifiable attribute based search over shared cloud data," in Proc. of the 35th International Conference on Data Engineering, Macao, China, 2019, 2028–2031.
[7] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. Shen, "Providing task allocation and secure deduplication for mobile crowd sensing via fog computing," IEEE Transactions on Dependable and Secure Computing, 17(3), 2018, 581–594.
[8] M. U. Arshad, A. Kundu, E. Bertino, A. Ghafoor, and C. Kundu, "Efficient and scalable integrity verification of data and query results for graph databases," IEEE Transactions on Knowledge and Data Engineering, 30(5), 2018, 866–879.
[9] Yang, J. Xu, J. Weng, J. Zhou, and D. S. Wong, "Lightweight and privacy-preserving delegatable proofs of storage with data dynamics in cloud storage," IEEE Transactions on Cloud Computing, 2018, DOI: 10.1109/TCC.2018.2851256.
[10] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. Shen, "Privacy-preserving smart parking navigation supporting efficient driving guidance retrieval," IEEE Transactions on Vehicular Technology, 67(7), 2018, 6504–6517.
[11] W. Zhang, Y. Lin, and Q. Gu, "Catch you if you misbehave: Ranked keyword search results verification in cloud computing," IEEETransactions on Cloud Computing, 6(1), 2018, 74–86.
[12] C. Guo, R. Zhuang, Y. Jie, K. Choo, and X. Tang, "Secure range search over encrypted uncertain iot outsourced data," IEEE Internet of Things Journal, 6(2), 2018, 1520–1529.
[13] Y. Wang, Q. Wu, B. Qin, W. Shi, R. Deng, and J. Hu, "Identity based data outsourcing with comprehensive auditing in clouds," IEEE Transactions on Information Forensics and Security, 12(4), 2017, 940–952.
[14] X. Yao, R. Zhang, Y. Zhang, and Y. Lin, "Verifiable social data outsourcing," in Proc. of IEEE Conference on Computer Communications, Atlanta, USA, 2017, 1-9
[15] Y. Zhang, C. Xu, X. Liang, H. Li, Y. Mu, and X. Zhang, "Efficient public verification of data integrity for cloud storage systems from in distinguishability obfuscation," IEEE Transactions on Information Forensics and Security, 12(3), 2017, 676–688.
[16] K. Yang, Q. Han, H. Li, K. Zheng, Z. Su, and X. Shen, "An efficient and fine-grained big data access control scheme with privacy preserving policy," IEEE Internet of Things Journal, 4(2), 2017, 563–571.
[17] J. Li, R. Ma, and H. Guan, "Tees: An efficient search scheme over encrypted data on mobile cloud," IEEE Transactions on CloudComputing, 5(1), 2017, 126–139.
[18] H. Tian, Y. Chen, C.-C. Chang, H. Jiang, Y. Huang, Y. Chen, and J. Liu, "Dynamic-hash-table based public auditing for secure cloud storage," IEEE Transactions on Services Computing, 10(5), 2017, 701–714.