e-ISSN: 2349-9745
p-ISSN: 2393-8161

Scientific Journal Impact Factor (SJIF): 1.711

International Journal of Modern Trends in Engineering and Research

www.ijmter.com

Survey on Privacy- Preserving Multi keyword Ranked Search over Encrypted Cloud data

MONIKA.S[1],RAMASAMY.S[2]

[1]*PG Scholar, Department of Computer Science and Engineering, Vivekananda College of Engineering for Women*

[2]*Assistant Professor, Department of Computer Science and Engineering, Vivekananda College of Engineering for Women*

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.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. We first propose a basic MRSE scheme using secure inner product computation, and then significantly improve it to meet different privacy requirements in two levels of threat models. The Incremental High Utility Pattern Transaction Frequency Tree (IHUPTF-Tree) is designed according to the transaction frequency (descending order) of items to obtain a compact tree.

By using high utility pattern the items can be arranged in an efficient manner. Tree structure is used to sort the items. Thus the items are sorted and frequent pattern is obtained. The frequent pattern items are retrieved from the database by using hybrid tree (H-Tree) structure. So the execution time becomes faster. Finally, the frequent pattern item that satisfies the threshold value is displayed.

Keywords: Cloud Computing, Keyword Search, Privacy Preserving, Ranked Search, Searchable Encryption

I. INTRODUCTION

1.1 CLOUD COMPUTING

Cloud computing is the use of resources that are delivered as a service over a network. The name comes from the use of a cloud-shaped symbol as an abstraction for the complex infrastructure it contains in system diagrams. Cloud computing entrusts remote services with a user's data, software and computation.The cloud providers manage the infrastructure and platforms on which the applications run.End users access cloud-based applications through web browser or a light-weight desktop or mobile app while the business software and user's data are stored on servers at a remote location.

Cloud computing allows enterprises to get their applications up and running faster, with improved manageability and less maintenance, and enables IT to more rapidly adjust resources to meet fluctuating and unpredictable business demand.Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that

can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

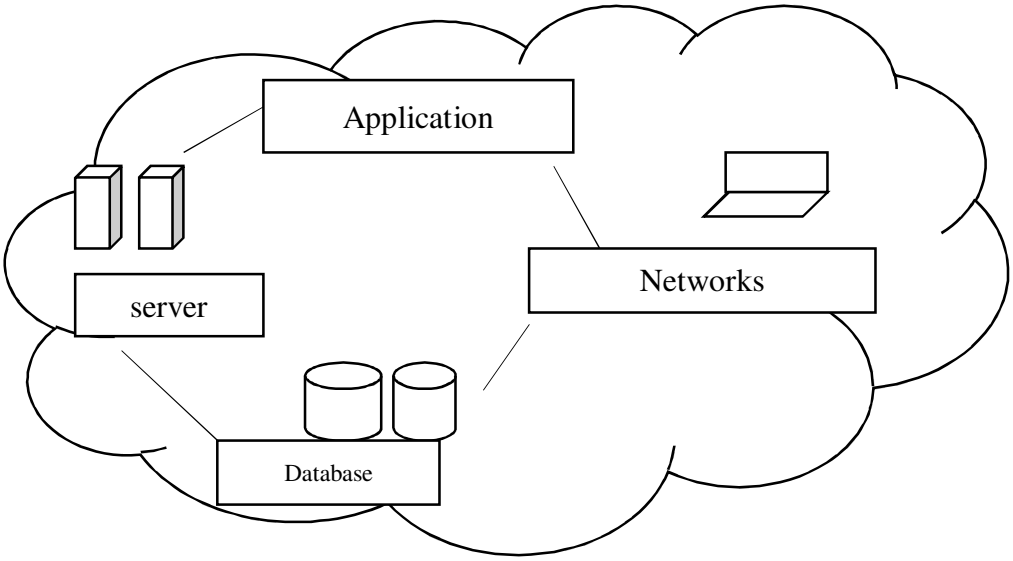

Fig 1: Architecture

1.2 CHARACTERISTICS

- Agility improves with users' ability to re-provision technological infrastructure resources.
- Device and location independenceenable users to access systems using a web browser regardless of their location or what device they are using (e.g., PC, mobile phone)
- Reliability is improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.
- Scalability and elasticity via dynamic ("on-demand") provisioning of resources on a fine-grained, self-service basis near real-time, without users having to engineer for peak loads.
- Security could improve due to centralization of data, increased security-focused resources, etc., but concerns can persist about loss of control over certain sensitive data, and the lack of security for stored kernels.Security is often as good as or better than other traditional systems
- Maintenance of cloud computing applications is easier, because they do not need to be installed on each user's computer and can be accessed from different places.

1.3 TYPES OF CLOUD

Public cloud

Public cloud applications, storage, and other resources are made available to the general public by a service provider. These services are free or offered on a pay-per-use model.

Community cloud

Community cloud shares infrastructure between several organizations from a specific

community with common concerns (security, compliance, jurisdiction, etc.), whether managed internally or by a third-party and hosted internally or externally.

Hybrid cloud

Hybrid cloud is a composition of two or more clouds (private, community or public) that remain unique entities but are bound together, offering the benefits of multiple deployment models.[4]

Private cloud

Private cloud is cloud infrastructure operated solely for a single organization, whether managed internally or by a third-party and hosted internally or externally.

II.LITERATURE SURVEY

2.1 Fuzzy Keyword Search over Encrypted Data in Cloud Computing

As cloud computing becomes prevalent, more and more sensitive information are being centralized into the cloud. Although traditional searchable encryption schemes allow a user to securely search over encrypted data through keywords and selectively retrieve files of interest, these techniques support only exact keyword search. For the first time we formalize and solve the problem of effective fuzzy keyword search over encrypted cloud data while maintaining keyword privacy. Fuzzy keyword search greatly enhances system usability by returning the matching files when users' searching inputs exactly match the predefined keywords or the closest possible matching files based on keyword similarity semantics, when exact match fails. In our solution, we exploit edit distance to quantify keywords similarity and develop two advanced techniques on constructing fuzzy keyword sets.

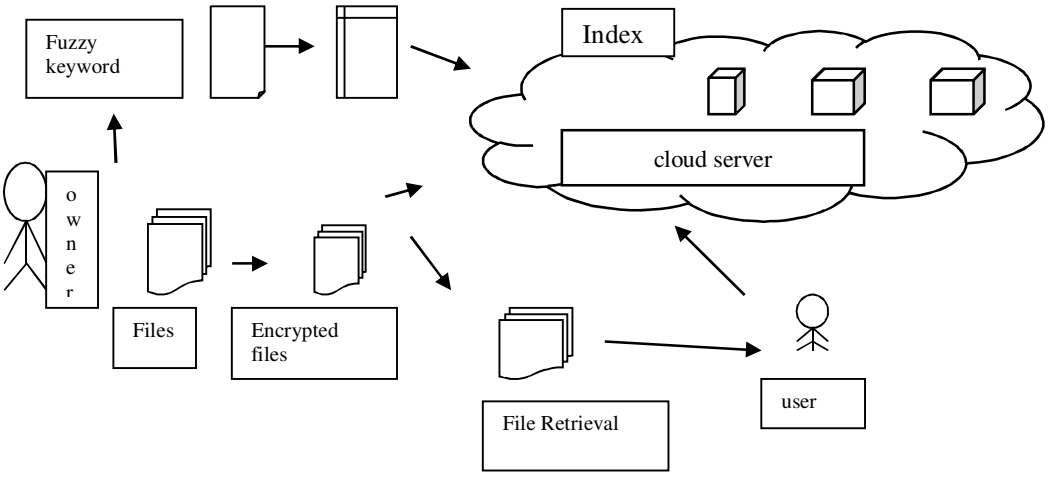

Fig 2: Architecture of Fuzzy Keyword Search

2.2 Achieving Secure, Scalable and Fine Grained Data Access Control in Cloud Computing

We use the technique of hybrid encryption to preserve data files, i.e. encrypt files using symmetric DEKs with KPABE. Using KP-ABE,this scheme is able to suddenly enjoy fine grained data access control and well organized operations such as file creation/deletion and new user grant.To resolve the challenging issue of user revocation,we combine the technique of proxy re-encryption with KP-ABE and reduce most of the problem in Cloud Servers.This scheme achieves this by keeping a partial copy of each user's secret key.When the data owner again specifies a certain set of attributes for the purpose of user revocation,he also produces corresponding proxy re-encryptionkeys and sends them to cloud servers.Cloud servers given these proxy re-encryption keys,can append user secret key componentsand re-encrypt data filesaccordingly without knowing the underlying data files.This enhancement releases the data ownerfrom the possible huge computation overheadon user revocation.The data owner also does notneed to always stay online.In order to save computation overheadof Cloud Servers on user revocation, the proposed scheme useslazy re-encryption techniqueand enable cloud servers to combine multiple successive secret keyupdate or file re-encryption operations into one.

2.3 Privacy Preserving Public Auditing for Secure Cloud Storage

Using Cloud Storage, users can remotely store their data and enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources, without the burden of local data storage and maintenance. Moreover, users should be ableto just use the cloud storage as if it is local, without worrying about the need to verify its integrity. Thus, enabling public auditability for cloud storage is of critical importance so that users can resort to a third party auditor (TPA) to check the integrity of outsourced data and be worry-free. To achieve privacy-preserving public auditing,we propose to uniquely integrate the homomorphic linear authenticator with random masking technique.In our protocol, the linear combination of sampled blocks in the server's response is masked with randomness generated the server. With random masking, the TPA no longer has all the necessary information to build up a correct group of linear equations and therefore cannot derive the user's data content, no matter how many linear combinations of the same set of file blocks can be collected. Our design makes use of a public key based HLA, to equip the auditing protocol with public auditability.

2.4 Towards Secure and Dependable Storage Services in Cloud Computing

Cong Wang, Qian Wang, and KuiRen discussed that Cloud Computing has been envisioned as the nextgeneration architecture of IT Enterprise. In contrast to traditional solutions, where the IT services are under proper physical, logical and personnel controls, Cloud Computing moves the application software and databases to the large data centers, where the management of the data and services may not be fully trustworthy. To ensure the correctness of users' data in the cloud, we propose an effective and flexible distributed scheme with two salient features, opposing to its predecessors. By utilizing the homomorphic token with distributed verification of erasure-coded data, our scheme achieves the integration of storage correctness insurance and data error localization.The new scheme further supports secure and efficient dynamic operations on data blocks, including: data update, delete and append.

By utilizing the homomorphic token with distributed verification of erasure-coded data, our scheme achieves the storage correctness insurance as well as data error localization: whenever data corruption has been detected during the storage correctness verification, our scheme can almost guarantee the simultaneous localization of data errors, i.e., the identification of the

misbehaving servers.

2.5 Fully Secure Functional Encryption: Attribute-Based Encryption and Inner Product Encryption

Fully secure functional encryption: Attribute-based encryption and (hierarchical) inner product encryption proposed a new way of viewing encryption which opens up a much larger world of possibilities for sharing encrypted data. In a functional encryption system, there is a functionality f(x, y) which determines what a user with secret key can learn from a ciphertext encrypted under x. In an IBE or HIBE system, keys and ciphertext are both associated with the same type of simple object.In an ABE system, keys and ciphertext are associated with more complex objects: attributes and access formulas. It uses a novel information-theoretic argument to adapt the dual system encryption methodology to the more complicated structure of ABE systems. We prove the security of our system from three static assumptions. Our ABE scheme supports arbitrary monotone access formulas. Security is proven under a non-interactive assumption whose size does not depend on the number of queries. The scheme is comparably efficient to existing selectively secure schemes. The key technique used to obtain these results is an elaborate combination of the dual system encryption methodology (adapted to the structure of inner product PE systems) and a new approach on bilinear pairings using the notion of dual pairing vector spaces (DPVS) proposed by Okamoto and Takashima.

2.6 Practical Techniques for Searches on Encrypted Data

The cryptographic schemes for the problem of searching on encrypted data provideprovable secrecy for encryption, query isolation for searching data, controlled searching.They also support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. The techniques provide provable secrecy for encryption, in the sense that the untrusted server cannot learn anything about the plaintext given only the ciphertext. It also provide controlled searching, so that the untrusted server cannot search for a word without the user's authorization and support hidden queries, so that the user may ask the untrusted server to search for a secret word without revealing the word to the server. Also support query isolation, meaning that the untrusted server learns nothing more than the search result about the plaintext.

2.7 Privacy-PreservingAudit and Extraction of Digital Contents

Mehul A. Shah Ram SwaminathanMary Baker says that growing number of online services, such as Google, Yahoo!, and Amazon, are starting to charge users for their storage. Customers often use these services to store valuable data such as email, family photos and videos, and disk backups. Today, a customer must entirely trust such external services to maintain the integrity of hosted data and return it intact. Unfortunately, no service is infallible. To make storage services accountable for data loss, we present protocols that allow a thirdparty auditor to periodically verify the data stored by a service and assist in returning the data intact to the customer. Most importantly, our protocols are privacy-preserving, in that they never reveal the data contents to the auditor. Our solution removes the burden of verification from the customer, alleviates both the customer's and storage service's fear of data leakage, and provides a method for independent

arbitration of data retention contracts.

2.8 Searchable Symmetric Encryption

Cloud computing has emerging as a promising pattern for data outsourcing and high quality data services. However, concerns of sensitive information on cloud potentially cause privacy problems. Searchable symmetric encryption (SSE) allows retrieval of encrypted data over cloud.Besides, in order to improve feasibility and save on the expense in the cloud paradigm, it is preferred to get the retrieval result with the most relevant files that match users' interest instead of all the files, which indicates that the files should be ranked in the order of relevance by users' interest and only the files with the highest relevance are sent back to users. A series of searchable symmetric encryption schemes have been proposed to enable search on ciphertext. Traditional SSE schemes enable users to securely retrieve the ciphertext, but these schemes support only Boolean keyword search, i.e., whether a keyword exists in a file or not, without considering the difference of relevance with the queried keyword of these files in the result. Preventing the cloud from involving in ranking and entrusting all the work to the user is a natural way to avoid information leakage. However, the limited computational power on the user side and the high computational overhead precludes information security.

2.9Secure Ranked Keyword Search over Encrypted Cloud Data

Cloud computing is a subscription-based service where the networked storage space and computer resources can be obtained. Cloud computing economically enables the paradigm of data service outsourcing. However, to protect data privacy, sensitive cloud data have to be encrypted before outsourced to the commercial public cloud, which makes effective data utilization. In the proposed system, the problem of effective secure ranked keyword search over encrypted cloud data is done. Ranked keyword search greatly enhances the system usability by returning the matching files in a ranked order. In Confidentiality-Preserving Rank-Ordered Search, when an authorized user remotely accesses the data to search and retrieve desired documents, the large size of the collections often makes it infeasible to ship all encrypted data to the user's side, and then perform decryption and search on the user's trusted computers. Therefore, new techniques are needed to encrypt and organize the data collections in a way as to allow the data centre to perform efficient search in an encrypted domain .Order-Preserving Symmetric Encryption(OPSE),is a deterministic encryption scheme whose encryption function preserves numerical ordering of the plaintexts. OPSE is the form of one-part codes, which are lists of plaintexts and the corresponding cipher texts, both of which are arranged in an alphabetical or numerical order so that a single copy is required for efficient encryption and decryption OPSE not only allows efficient range queries, but also allows indexing and query processing to be done exactly and is efficient for unencrypted data.

III.CONCLUSION

We have defined and solved the problem of multi-keyword ranked search over encryptedcloud data, and established a variety of privacy requirements.Among various multi-keyword semantics, we choose theefficient similarity measure of "coordinate matching," i.e., asmany matches as possible, to effectively capture therelevance of outsourced documents to the

query keywords,and use "inner product similarity" to quantitativelyevaluate such similarity measure. For meeting the challengeof supporting multi-keyword semantic without privacybreaches, we propose a basic idea of MRSE using secureinner product computation. Then, we give two improvedMRSE schemes to achieve various stringent privacy requirementsin two different threat models. We also investigatesome further enhancements of our ranked search mechanism,including supporting more search semantics, i.e., TF _IDF, and dynamic data operations.Thorough analysisinvestigating privacy and efficiency guarantees of proposedschemes is given, and experiments on the real-world dataset show our proposed schemes introduce low overhead onboth computation and communication.

REFERENCES

[1] L.M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner, "ABreak in the Clouds: Towards a Cloud Definition," ACM SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 50-55, 2009.
[2] D. Song, D. Wagner, and A. Perrig, "Practical Techniques forSearches on Encrypted Data," Proc. IEEE Symp. Security andPrivacy, 2000.
[3] C. Wang, Q. Wang, K. Ren, N. Cao, and W. Lou, "Toward Secureand Dependable Storage Services in Cloud Computing," IEEETrans. Services Computing, vol. 5, no. 2, pp. 220-232, Apr.-2013.
[4] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-PreservingPublic Auditing for Data Storage Security in Cloud Computing,"
Proc. IEEE INFOCOM, 2010.
[5] C. Wang, N. Cao, K. Ren, and W. Lou, "Enabling Secure andEfficient Ranked Keyword Search over Outsourced Cloud Data,"IEEE Trans. Parallel and Distributed Systems, vol. 23, no. 8, pp. 1467-
1479, Aug. 2012.
[5] S. Kamara and K. Lauter, "Cryptographic Cloud Storage," Proc.14th Int'l Conf. Financial Cryptograpy and Data Security, Jan. 2010.
[6] N. Cao, Z. Yang, C. Wang, K. Ren, and W. Lou, "PrivacypreservingQuery over Encrypted Graph-Structured Data in Cloud
Computing," Proc. Distributed Computing Systems (ICDCS),pp. 393-402, June, 2011.
[7] E.-J. Goh, "Secure Indexes," Cryptology ePrintArchive,http://eprint.iacr.org/2003/216. 2003.
[8] Y.-C. Chang and M. Mitzenmacher, "Privacy Preserving KeywordSearches on Remote Encrypted Data," Proc. Third Int'l Conf.Applied Cryptography and Network Security, 2005.
[9] R. Curtmola, J.A. Garay, S. Kamara, and R. Ostrovsky, "SearchableSymmetric Encryption: Improved Definitions and Efficient Constructions,"Proc. 13th ACM Conf. Computer and Comm. Security(CCS '06), 2006.
[10] D. Boneh, G.D. Crescenzo, R. Ostrovsky, and G. Persiano, "PublicKey Encryption with Keyword Search," Proc. Int'l Conf. Theory andApplications of Cryptographic Techniques (EUROCRYPT), 2004.
[11] J. Li, Q. Wang, C. Wang, N. Cao, K. Ren, and W. Lou, "FuzzyKeyword Search Over Encrypted Data in Cloud Computing,"
Proc. IEEE INFOCOM, Mar. 2010.
[12] D. Boneh, E. Kushilevitz, R. Ostrovsky, and W.E.S. III, "Public KeyEncryption That Allows PIR Queries," Proc. 27th Ann. Int'l
Cryptology Conf. Advances in Cryptology (CRYPTO '07), 2007.
[13] P. Golle, J. Staddon, and B. Waters, "Secure Conjunctive Keyword
Search over Encrypted Data," Proc. Applied Cryptography andNetwork Security, pp. 31-45, 2004.
[14] L. Ballard, S. Kamara, and F. Monrose, "Achieving Efficient Conjunctive Keyword Searches over Encrypted Data," Proc.
Seventh Int'l Conf. Information and Comm. Security (ICICS '05),2005.
[15] D. Boneh and B. Waters, "Conjunctive, Subset, and Range Querieson Encrypted Data," Proc. Fourth Conf. Theory Cryptography (TCC),pp. 535-554, 2007.
[16] R. Brinkman, "Searching in Encrypted Data," PhD thesis, Univ. ofTwente, 2007.
[17] Y. Hwang and P. Lee, "Public Key Encryption with ConjunctiveKeyword Search and Its Extension to a Multi-User System,"Pairing, vol. 4575, pp. 2-22, 2007.
[18] M.A. Shah, R. Swaminathan, and M. Baker, "Privacy-Preserving Audit and Extraction of Digital Contents," Cryptology ePrint Archive, Report 2008/186, 2008.

