An Efficient Technique for the cloud Data Retrieval using IBE

Parul Khatri, Asst. Prof. Sandeep Kumar

Abstract— Since Cloud computing is a new fields having more research work to be done specially in the field of security. Cloud Computing enables various users to send the data over internet which is then stored at data centers, but there is less chance that data is secure and can't be access by the un-authorized users or chances of data loss. Hence Security is an important concern in the cloud computing. Although there are various techniques implemented for the security of cloud data so that it can be access by the un-authorized users and privacy is maintained on user's data. Here in this paper a new and efficient technique is implemented for the security of cloud storage data and quick retrieval of cloud data.

Index Terms— Vm, PAAS, SAAS, K-retrieval, ECC, AES, ECIES, DES.

1 INTRODUCTION

Clouds can be explained as pools of virtualized resources that can be easily used and accessed [1]. For optimum resource utilization the resources in cloud can be reconfigured dynamically. With the help of strong cloud architectures its mass computing and storage centers organizations and individuals are benefited while utilizing them. Cloud computing basically contains virtualization, on-demand deployment, Internet delivery of services, open source software etc. [2].

With the help of internet and central remote servers cloud computing maintains data and applications. Cloud Computing provides efficient computing by centralizing storage, memory, processing and bandwidth promising lower costs, rapid scaling, easier maintenance, service availability. The main focus needs upon the data security and privacy. Services provided by cloud computing are [3]:

- Services to large number of distinct end users in opposition to bulk
- Developers are capable of running applications on a separate computing platform with physical infrastructure, job scheduling, user authentication, base software environment etc. and do not need to implement platform by themselves.

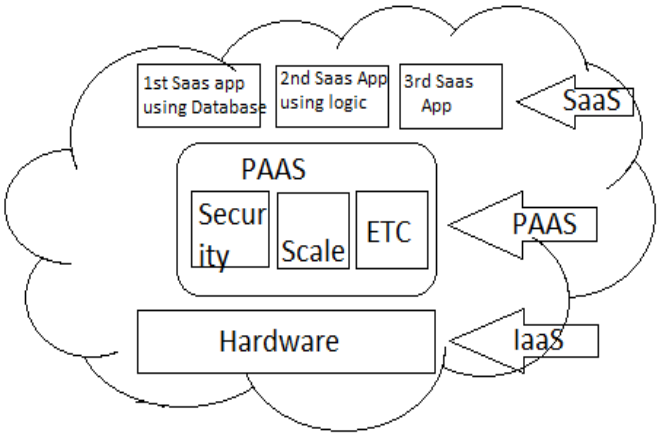

Figure 1: Cloud Computing Services

CLOUD SECURITY

Security is an important issue during the transmission of data from sender to receiver whether in wired network or in wireless network or in cloud computing. Cloud Computing provides transmission of data from user to data centre through brokers over internet. Hence security of data is important such that it prevents from various attacks in cloud computing [4].

TPA provides one of the efficient and secure techniques to provide prevention from various attacks. Here public verifiability is needed to ensure the data access from the user is authorized or not [5]. Privacy preserving with public auditing also enables the data to be made secure with homomorphic non-linear authentication and verification [6]. Verifiability of two types private Verifiability and public Verifiability [8].

TOP-K RETRIEVAL

Page ranking is a concept of providing the best retrieval of web pages according to their ranking. The data from various users in cloud computing can be stored in storage panel according to their identities or keywords [10].

With information needs emerging beyond a simple exact match paradigm, databases and information systems have since long catered for extended retrieval paradigms like top-k retrieval or skyline queries. Since the top-k paradigm has been first introduced into the area of database systems a large number of different algorithms have been proposed [11], [12]. Algorithms for top-k retrieval in databases generally try to minimize the number of database objects that have to be accessed before being able to return a correct result set of the k best matching objects. The problem how to define correct retrieval gets even worse, if top-k queries have to be answered. Besides the difficulties with the volatility of the P2P network, also the heterogeneity of the peers plays an important part, since each peer only knows its local objects and different peers may also feature different scoring or retrieval strategies [13].

ENCRYPTION TECHNIQUES

Some amount of data security can be achieved through the encryption of the data.
Some encryption techniques are:

ECIES: The Elliptic Curve Integrated Encryption Scheme is a public-key encryption scheme based on ECC. It is designed to be semantically secure in the presence of an adversary capable of launching chosen-plaintext and chosen-cipher text attacks.
The advantage of the ECIES is that on one hand it is meanwhile quite well investigated and thus considered secure while on the other hand just a very short bit length is needed as compared to other asymmetric systems.
ECC: Elliptical curve cryptography is a public key encryption technique based on elliptic curve theory that can be used to make faster, less significant and more competent cryptographic keys. ECC produces keys during the properties of the elliptic curve equation as an alternative of the conventional method of creation as the produce of very huge prime numbers. The technology can be used in coincidence with most public key encryption methods, such as RSA, and Diffie-Hellman. According to some examiners, ECC can give way a stage of security with a 164-bit key that other schemes require a 1,024-bit key to accomplish.
AES: It is stand for Advanced Encryption Standard. It is a specification of the electronic data encryption. The Advanced Encryption Standard comprises three block ciphers, AES-128, AES-192 and AES-256. AES as a fixed block size of 128 bits and a key size of 128, 192, or 256 bits. The block-size has a maximum of 56 bits, but the key-size has no theoretical maximum. The cipher uses number of encryption rounds which converts plain text to cipher text.
DES: It stands for Data encryption standard. It is a widely-used method of data encryption with the help of private or secrete key. DES uses 56-bit key to each 64-bit block of data. It can run in various modes and involves 16 rounds or operations. Although this is considered "strong" encryption, many companies used 'triple DES' that uses three keys in succession.

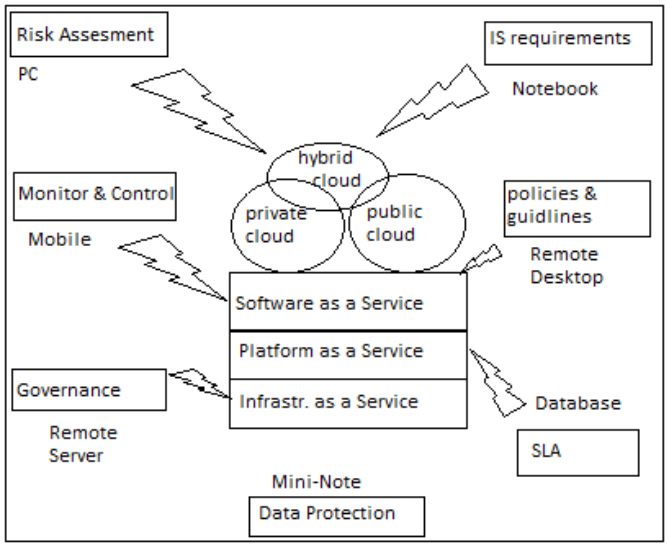

Figure 2: Scenario of Retrieval of Encrypted Cloud Data & models [1].

2 Literature Survey

Some common approaches are also discussed here that work efficiently in field of cloud storage & security, virtualization, multi keyword top-k retrieval, ECIES and other encryption techniques.
Jiadi Yu et al [1] uses the concept of multi-keyword retrieval of data from cloud computing. The idea is to encrypt the data based on the keyword of the data and stored in storage panel with their respective encrypted data and keyword. So that when a keyword is search a SSE technique is used to search the keyword and the matched keywords gets the encrypted data and can be decrypted further.

Swathi Sambaing [4] proposed publicly auditable cloud storage providers where data owners can rely on third party auditor to verify the data integrity of sourced data to make sure security.

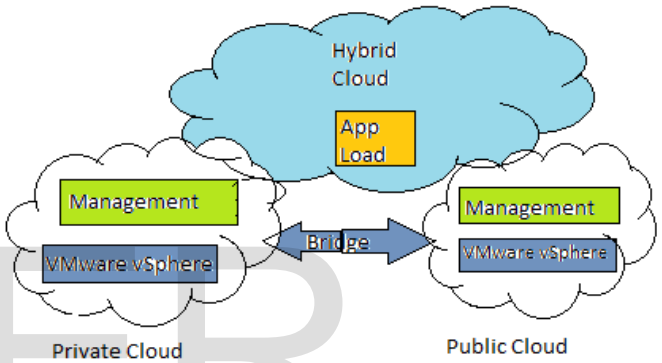

Figure 3: Audit system architecture for cloud computing [4].

Q. Wang et al [5] explored the problem of providing simultaneous public auditability and data dynamics for remote data integrity check in Cloud Computing. In view of the key role of public auditability and data dynamics for cloud data storage they propose an efficient construction for the seamless integration of these two components in the protocol design.

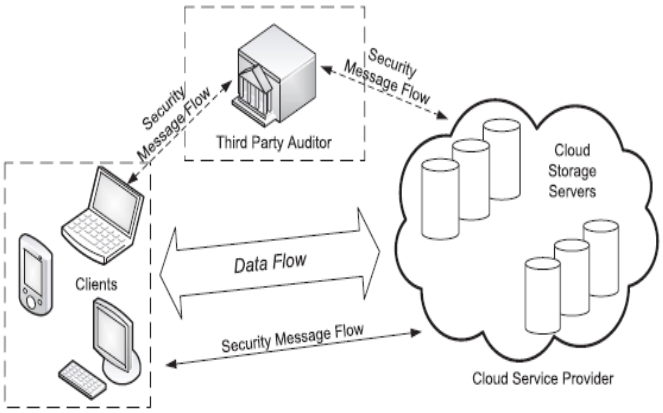

Figure 4: Cloud Data Storage Architecture [5].

They [5] assume that TPA is unbiased while the server is untrusted. For application purposes, the clients may interact with

the cloud servers via CSP to access or retrieve their pre stored data. More prominently in realistic scenarios, the client may recurrently perform block-level operations on the data files. The most familiar forms of these operations are modification, insertion, and deletion. To effectively support public auditability without having to retrieve the data blocks themselves, they resort to the homomorphic authenticator technique. The naive way of realizing data integrity verification is to make the hashes of the original data blocks as the leaves in MHT, so the data integrity verification can be conducted without tag authentication and signature aggregation steps. So they adopt the block less approach, and authenticate the block tags instead of original data blocks in the verification process. To achieve efficient data dynamics a new and efficient technique is implemented.

Srinivas, D. proposed a new technique in which the burden of cloud user from the tedious and possibly pricey auditing task, but also alleviates the users' terror of their outsourced data security. Taking into account TPA may concurrently handle multiple audit sessions from dissimilar users for their outsourced data files, he further extend this privacy-preserving public auditing protocol into a multi-user scenario, where the TPA can perform multiple auditing tasks in a batch manner for better effectiveness. Extensive examination shows that this scheme is almost certainly secure and highly efficient [6].

Zhu, Yan et al [7] suggested efficient provable data possession for hybrid clouds. They focused on the construction of PDP scheme for hybrid clouds, supporting privacy protection and dynamic scalability. They first provide an effective construction of Cooperative Provable Data Possession (CPDP) using Homomorphic Verifiable Responses (HVR) and Hash Index Hierarchy (HIH). This construction uses homomorphic property, such that the responses of the client's challenge computed from multiple CSPs can be combined into a single response as the final result of hybrid clouds. By using this mechanism, the clients can be convinced of data possession without knowing what machines or in which geographical locations their files reside. More prominently a new hash index hierarchy is proposed for the clients to seamlessly store and manage the resources in hybrid clouds. Their experimental results also validate the effectiveness of their construction.

In 2009 Qian Wang et al [8] introduced a new scheme which gives remote data integrity and verifiability means dynamic data operations. The method initially identifies the troubles and potential security problems of direct extensions with fully dynamic data updates. It achieves efficient data dynamics and improves the Retrieve ability model by manipulating the classic Merkle Hash Tree (MHT) construction used for block tag validation. It is extremely proficient and secure technique [8].

Michael Armbrust et al [9] presented a survey on cloud computing. They defined basic terminology of cloud. They also compare cloud with other related technologies. They also try to identifying the top technical and non-technical obstacles and opportunities of cloud computing. Virtualization is primary security mechanism of cloud computing. Multiple virtual machines (VMs) can share CPUs and main memory surprisingly well in cloud computing. Virtualization is essential to improve architectures and operating systems to efficiently virtualized interrupts and I/O channels [9].

The problem of effective yet secure ranked keyword search over encrypted cloud data is solved by Cong Wang et al [14]. Ranked search greatly enhances system usability by returning the matching files in a ranked order regarding to certain relevance criteria, thus making one step closer towards practical deployment of privacy-preserving data hosting services in Cloud Computing. They primarily given a straightforward yet ideal construction of ranked keyword search under the state-of-the-art searchable symmetric encryption (SSE) security definition, and demonstrate its inefficiency. To accomplish more practical performance, then they suggested a definition for ranked searchable symmetric encryption, and given an efficient design by properly utilizing the existing cryptographic primitive, order-preserving symmetric encryption (OPSE). Thorough analysis shows that this solution enjoys "as-strong-as-possible" security guarantee compared to previous SSE schemes, while correctly realizing the goal of ranked keyword search. Extensive experimental results demonstrate the efficiency of the proposed solution [14].

Cong Wang et al [15] recommended a secure cloud storage system supporting privacy-preserving public auditing. This is among the first few ones to support privacy-preserving public auditing in cloud computing, with a focus on data storage. Besides, with the prevalence of cloud computing, a foreseeable increase of auditing tasks from different users may be delegated to TPA. Their preliminary experiment conducted on Amazon EC2 instance further demonstrates the fast performance of their design on both the cloud and the auditor side. They leaved the full-fledged implementation of the mechanism on commercial public cloud as an important future extension that is expected to robustly cope with very large scale data and thus encourage users to adopt cloud storage services more confidently [15].

Ning Caoy et al [16] define and solve the problem of multi-keyword ranked search over encrypted cloud data (MRSE) while preserving strict system-wise privacy in the cloud computing paradigm. To meet the challenge of supporting such multi-keyword semantic without privacy breaches, they suggested a basic idea for the MRSE using secure inner product computation, which is adapted from a secure k-nearest neighbor (kNN) technique [17], and then give two significantly improved MRSE schemes in a step-by-step manner to achieve various stringent privacy requirements in two threat models with increased attack capabilities.

They [16] defined the framework of multi-keyword ranked search over encrypted cloud data (MRSE) and establish various strict system-wise privacy requirements for such a secure cloud data utilization system. For easy presentation, operations on the data documents are not shown in the framework since the data owner could easily employ the traditional symmetric key cryptography to encrypt and then outsource data. The representative privacy guarantee in the related literature, such as searchable encryption, is that the server should learn nothing but search results. With this general privacy description, they explored and established a set of strict privacy requirements specifically for the MRSE framework.

Hussain Abo Surrah aimed to provide searching a file over cloud environment using multiple keywords representing the file with various probable situations [18]. The target is to provide the security to its maximum extent by including encryption and decryption methods. Authorization of the users directly by the administrators allows the files involved to transfer more securely. Encryption and decryption of both file name and file which uses asymmetric and symmetric key algorithms respectively. The secret key is generated for each user to prevent any other user to misuse the file. The data that are stored in the cloud has to be protected completely from any attack that is caused both by external and internal attackers. Most of the internal attacks are used by the cloud providers by using similarity relevance and analyzing the statistical leakage. Based on the usage of the file over ranked manner, it is easy to get all the details of the most used files through probability prediction. This kind of data leakage should be completely avoided and maximum protection to the data is given. The solution suggests the same by applying some new concepts to increase the data security [18].

3 Proposed Methodology

The proposed methodology works on the following four phases:
1. Setup
2. Key Generation
3. Encryption
4. Decryption

Setup phase & Key Generation
During the setup of the proposed methodology the parameters of the cloud needs to be initialize such as users, brokers and data centres as well as the physical characteristics of the cloud also needs to be setup.
Since Elliptic Curve Cryptography is used here for the generation of public and private keys, hence the basis elliptic curve equation of the form:

$$y^3 = ax^3 + bx + c, where\ 4a^2 + 27b^2 \neq 0.$$

Here sender and receiver need to select a private random point on the elliptic curve and a common base point G. From the generated private and Base point public key is generated using,

$$x.G = y$$

Where, 'x' is the private key and G is the common base point and 'y' is the public key.

4.2.2 Encryption
For the encryption of the message with a keyword 'K' using public key that can be derived from string 'str'. For every string that contains a keyword and data and time known as 'str'. First of all generate a public key for the known bit string and applying identity based encryption to obtain ciphertext 'C'.

Decryption
The receiver for the decryption of the ciphertext 'C' uses his private key to generate original message m'.

Algorithm

1. Setup the cloud environment with a number of users and data centres and brokers having their individual physical characteristics.
2. User 'Ui' when sends the data to the data centre will generate a keyword and create a string 'str'.
3. User 'Ui' using his public key encrypt the data and send to the storage repository in the form of tupple (keyword, cipher text).
4. User 'Ui' also allots a unique id and password for the receiver for the access of the data.
5. Te receiver needs to authenticate first for the data to access.
6. After authentication receiver 'R' sends query in form of keyword to the central authority where on the basis of keyword the queries are fetched with the match keyword.
7. Receiver accesses the data in encrypted form and performs decryption using private key.
8. Receiver also verifies the message is valid or not using Message Authentication Code.

4 Result Analysis

The figure shown below is the analysis and comparison of existing and proposed work. The analysis is done on the basis of number of keywords and computational time to access these keywords.

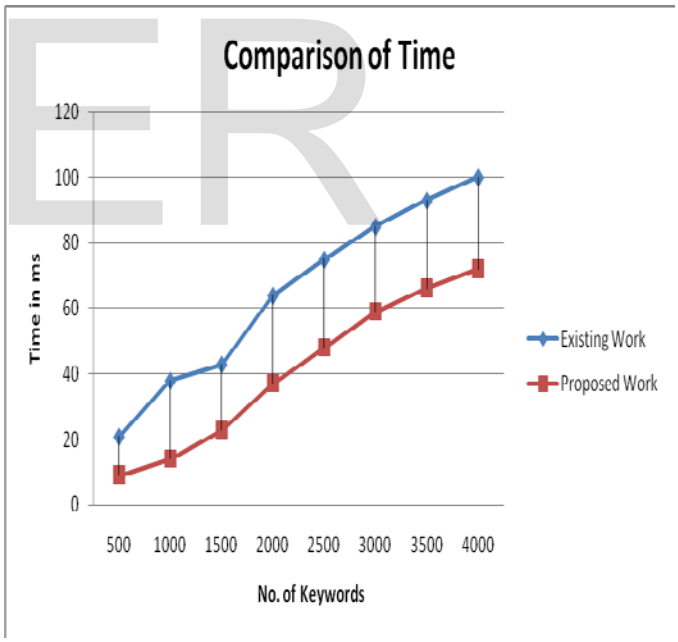

Figure 5 Time Vs No. of Keywords comparisons

The figure shown below is the analysis and comparison of existing and proposed work. The analysis is done on the basis of number of files and computational time to access these keywords.

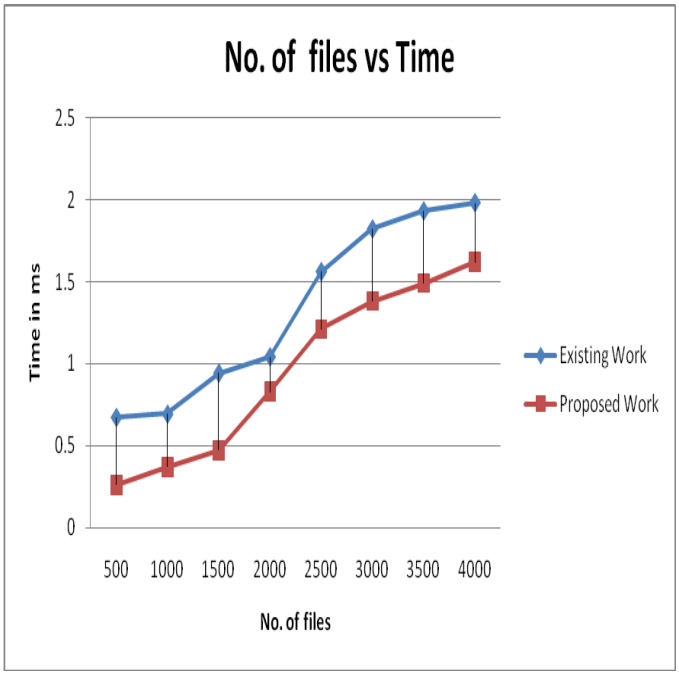

Table 6 Comparison of Time on No. of files

5 Conclusion

The proposed methodology implemented here for the retrieval of encrypted data using ECIES is efficient in terms of retrieval rate and security and storage cost and computational time. The existing technique implemented for the retrieval of encrypted data suffers from various attacks and the proposed scheme guarantees data privacy. According to the efficiency evaluation of the proposed scheme over a real data set, extensive experimental results demonstrate that our scheme ensures practical efficiency, but the technique implemented here prevents from the above issues hence performance is better as compared to the existing technique.

References

[1] Yu, Jiadi, Peng Lu, Yanmin Zhu, Guangtao Xue, and Minglu Li. "Towards Secure Multi-Keyword Top-k Retrieval over Encrypted Cloud Data," IEEE transactions on dependable and secure computing, vol. 10, no. 4, pp. 239- 250, July/August 2013.

[2] Pankaj Arora, Rubal Chaudhry Wadhawan and Er. Satinder Pal Ahuja "Cloud Computing Security Issues in Infrastructure as a Service", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277-128X, vol. 2, issue 1, Jan. 2012.

[3] Song, Dawn, Elaine Shi, Ian Fischer, and Umesh Shankar. "Cloud data protection for the masses", In IEEE Computer, vol. 45, no. 1, pp. 39-45, 2012.

[4] Swathi Sambangi "Cloud Data Storage Services Considering Public Audit for Security", Global Journal of Computer Science and Technology Cloud and Distributed, ISSN: 0975-4172, Vol. 13, Issue 1, pp. 1 – 6, 2013.

[5] 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 and Distributed Systems, vol. 22, no. 5, pp. 847-859, May 2011.

[6] Srinivas, D. "Privacy-Preserving Public Auditing In Cloud Storage Security." International Journal of computer science and Information Technologies, ISSN: 0975-9646, vol. 2, no. 6, pp.2691-2693, 011.

[7] Zhu, Yan, Huaixi Wang, Zexing Hu, Gail-Joon Ahn, Hongxin Hu, and Stephen S. Yau. "Efficient provable data possession for hybrid clouds." In Proceedings of the 17th ACM conference on Computer and communications security, pp. 756-758. ACM, 2010.

[8] Qian Wang, Cong Wang, Jin Li1, Kui Ren, and Wenjing Lou, "Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing" Proceedings of the 14th European conference on Research in computer security(ESORICS'09), pp. 355-370, 2009.

[9] Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee et al. "A view of cloud computing." Communications of the ACM, vol. 53, no. 4, pp. 50-58, 2010.

[10] Chaudhuri, Surajit, and Luis Gravano. "Evaluating top-k selection queries." In VLDB, vol. 99, pp. 397-410. 1999.

[11] Fagin, Ronald, Amnon Lotem, and Moni Naor. "Optimal aggregation algorithms for middleware." Journal of Computer and System Sciences, vol. 66, no. 4, pp. 614-656, 2003.

[12] Balke, Wolf-Tilo, and Werner Kießling. "Optimizing multi-feature queries for image databases." In Proceedings of the International Conference on Very Large Databases, 2000.

[13] Balke, W-T., Wolfgang Nejdl, Wolf Siberski, and Uwe Thaden. "Progressive distributed top-k retrieval in peer-to-peer networks." In Proceedings of IEEE 21st International Conference on Data Engineering (ICDE 2005), pp. 174-185, 2005.

[14] Wang, Cong, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou. "Secure ranked keyword search over encrypted cloud data." In Distributed Computing Systems (ICDCS), 2010 IEEE 30th International Conference on, pp. 253-262. IEEE, 2010.

[15] Cong Wang, Sherman S.M. Chow, Qian Wang, Kui Ren, and Wenjing Lou, "Privacy-Preserving Public Auditing for Secure Cloud Storage" IEEE TRANSACTIONS ON COMPUTERS, VOL. 62, NO. 2, FEBRUARY 2013.

[16] Cao, Ning, Cong Wang, Ming Li, Kui Ren, and Wenjing Lou. "Privacy-preserving multi-keyword ranked search over encrypted cloud data." Parallel and Distributed Systems, IEEE Transactions on 25, no. 1 (2014): 222-233.

[17] Yousef Elmehdwi, Bharath K. Samanthula and Wei Jiang," Secure *k*-Nearest Neighbor Query over Encrypted Data in Outsourced Environments", 2013.

[18] Wong, Wai Kit, David Wai-lok Cheung, Ben Kao, and Nikos Mamoulis. "Secure kNN computation on encrypted databases." In Proceedings of the 2009 ACM SIGMOD International Conference on Management of data, pp. 139-152, 2009.