Secure Multi-Keyword Top K Similarity Search Using Asymmetric Encryption Over Encrypted Cloud Data

Khushpreet Kaur and Meenakshi Bansal

Abstract: Cloud computing provides the facility for both massive data storage and management. A cloud has large capacity to store data from number of users and individual and offers an easy option for the fetching of data anytime. Nowadays a number of organizations are shifting their fields towards cloud for the data storage and retrieval and the reason for so is the complex data management at economical cost along with great flexibility. But then comes the issue of security and hence cloud offers the security techniques that encrypt the data before outsourcing to the cloud server and in turn it also reduces the data utility. In the existing systems, the Symmetric Key Encryption is made to use but is less secure as compared to Asymmetric Key Encryption. Furthermore, the Symmetric Key Encryption offers a limited numbers of cluster that are used to access the secured data. Hence, to improve the query efficiency, data security and data utility the scheme named Group Multi-Keyword Top-K Search is used. In this paper, we focus on investigating the multi-keyword top-k search issue for big data encryption against privacy openings and trying to identify a secure but effective solution. Unambiguously, we focus on constructing a special tree-based index structure for the privacy of query data and designing a random traversal algorithm that can make even the same query to generate different visiting paths on the same index, while at the same time maintaining the accuracy of queries that are unchanged under the stronger privacy. This system is dependent on the hierarchical cluster and each cluster is then divided into sub-clusters till a minimum possible size of cluster is reached. Unlike the existing system, the proposed system uses the Asymmetric Encryption to fortify the security of data on the cloud. Main emphasis of this study is to fetch the outsourced encrypted data from the cloud by assigning the attribute key to the data.

Keywords: Group multi-keyword search, Asymmetric SE scheme, Cloud computing, Data encryption, random traversal, multi-keyword top-k search.

1 INTRODUCTION

CLOUD computing has emerged as a hottest research field in both IT industries and research communities. It became so popular because of its number of prominent features such as high scalability and pay-as-you-go mode. This feature of cloud computing have enabled consumers to purchase powerful computing resources such as services according to their needs without worrying about the complexity of hardware platform management. Today, more and more companies dealing with big data and individuals are outsourcing information and uploading their products to cloud servers for easy data management, well-organized data mining and request processing tasks. However, the privacy issue of outsourced data needs to be taken into consideration. The data sets which are uploaded on clouds also contain personal and sensitive information including emails, electronic health records and financial transaction reports. This data may also include any confidential or sensitive information such as business secret data, credit card details, banking transactions, etc. The discloser of such confidential information to any unauthorized user may lead to harmful consequences. Therefore this data need to be protected. There is a high security obligation when the confidential data is transmitted over an untrusted network. Data encryption has been widely used for data privacy and data maintenance in data sharing scenarios. Variety of data encryption frameworks have been developed and are used before outsourcing the data to cloud. We therefore need to establish a highly secure system for translating confidential data into some unreadable form to make it impossible for the hacker to get accurate data.

1.1 Security Techniques & Issues

The ultimate goal of cryptography is to prevent surveillance and obtain the original data. It enables that only authorized users should access the exact data without any modification. Cryptography techniques ensure the confidential of data from unauthorized users so that data remains safe and secure. The complete process of converting the data to a unreadable form and returning to get the original data back is called encryption and decryption respectively. These techniques are generally classified into two categories.

a) Transposition technique
b) Substitution technique

Cryptography is classified into two parts as Symmetric and Asymmetric cryptographic method, depending on the type of keys used for cryptographic operation on the plaintext.

Symmetric Cryptography: In this technique, the message is encrypted and decrypted using same secret key. Sender and receiver share this secret key for communication. As per the procedure, the sender must generate a secret key and then use that secret key to encrypt the plaintext and send it to the recipient. Receiver use the same secret key exchanged between these two parties to decode the received cipher text and get back the original information. Symmetric cryptographic operation is more effective and faster than asymmetric cryptographic operation. Figure 1 shows the cryptographic process of the symmetric key. Asymmetric Cryptography: In this technique, two keys are used one as public key and other as private key to encrypt and decrypt messages. Public key is known to everyone and used for plaintext authentication, while private key is known only to the individual who can perform the decryption to recover the original data from the obtained cipher text. Here, some mathematical mean correlates both the public and private keys.

- *Kushpreet Kaur is currently pursuing masters degree program in Computer Science and Engineering in Yadawindra College of Engineering, Punjabi University, Talwandi Sabo, PH-9478906966. E-mail: kushiwander16@gmail.com.*
- *Dr. Meenakshi Bansal is a Supervisor currently working Assistant Professor in Computer Science and Engineering in Yadawindra College of Engineering, Punjabi University, Talwandi , PH-8146800408. E-mail: author_ermeenu10@gmail.com*

1.2 Security Issues

- Confidentiality: Security necessity for which the intended client must interpret the message correctly. To do this, it is important to avoid unauthorized access and use.
- Authenticity Security requirement in which the network nodes should deliver its service without failure.
- Integrity Security condition which ensures that the data send by the sender is same as received by the receiver.
- Authenticity This security prerequisite verifies the identity of the node.
- Authorization Security requirement to ensure that only licensed sensors are needed to disseminate information.
- Nonrepudiation Security requirement ensures that the sender could not deny the sending of the message.
- Freshness The safety requirement concerns the maintenance and dissemination by sensor nodes of up-to-date information.

2 LITERATURE SURVEY

Literature review goes beyond knowledge quest and involves defining the connections between the literature and our research field. The cryptographic scheme was described by Song et al. (2000) to solve the problem of searching for encrypted data and to provide proof of security for the resulting crypto system. They use other security methods such as being proved secure. Seth et al. (2011) made the comparison between three algorithms, DES, AES and RSA considering the parameters like query time, data usages and output computer memory unit. It had been over that RSA consumes average time to encrypt the data and memory usage is additionally terribly high however output computer memory unit is least just in case of RSA algorithmic rule. Liang et al. (2013) comprehensive mobile cloud computing study presented could be a promising technique for moving database and computer system modules data from individual devices to geographically dispersed cloud service architecture. Multiple cloud domains consist of a general mobile cloud computing, and each domain manages some of the cloud system resources, such as the Central Process Unit, memory and storage, etc. How to manage cloud resources efficiently across growing cloud domains is essential to the continuous delivery of mobile cloud services. Cao et al. (2013) presented the detailed study of transfer and upload the encrypted data on cloud server. Because of the introduction of cloud servers, data holders were allowed to migrate their integrated data executive systems from local sites to the corporate public cloud for large donation and financial savings. Suggest a simple idea for MRSE in this project based on safe internal consumer computation and then offer two dramatically improved MRSE schemes to achieve different inflexible privacy criteria in two different risk models and enhance the information search system experience. Jung et al. (2013) had conducted the research on multiple parties share the secure data on cloud server. multiple parties' information aggregation to an entrusted aggregator while not revealing every individual's in private closely-held information, or to alter multiple parties to put together combination their information whereas protective privacy. Cong Wangy, Kui Reny, Shucheng Yux, and Karthik Mahendra Raje Urs in 2013 presented Usable and privacy-assured check for similarities over outsourced cloud data. In this way, it is getting more pervasive than some other time in late memory for data proprietors to outsource the psyche boggling data organization structures from close-by machines to cloud for the titanic versatility and cost reserves. However, before outsourcing, confidential information should be mixed by software proprietors in order to maintain data protection and battle unconstrained in the cloud or past. Given the generous number of on-ask requests for data customers and the enormous amount of outsourced data records in the cloud, the issue is particularly challenging, as it is incredibly difficult to meet the realistic requirements of execution, structure comfort, and anomalous state customers looking for experiences in the same way. We formally show the insurance sparing affirmation of the proposed part under exhaustive security treatment. To show the comprehensive proclamation of our instrument and further upgrade the application go, we moreover exhibit our new improvement ordinarily supports fleecy interest, a some time ago considered idea directing just toward bear linguistic oversights and depiction anomalies in the customer looking for input. The expansive tests on Amazon cloud organize with honest to goodness instructive gathering furthermore show the authenticity and presence of mind of the proposed framework. In this paper, roused by finishing sensible structure. Comfort and strange state customer seeking information, we investigate outsourced cloud data on the topic of stable and profitable similarity. With change evacuate as the similarity metric, at first our segment layout misused a smothering technique to construct a boundary beneficial resemblance catchphrase collection from an amassing document. Using that watchword set as an introduce, we by then propose another picture based trie-cross looking segment, and show it viably finishes the described likeness look for with enduring request time multifaceted nature. Li et al. (2014) presented a detailed study of cloud computing infrastructure may be a promising new technology, promoting the event of large-scale storage, processing and distribution of information. Tang et al. (2016) conducted the research on the cloud server to secure the shared data. Due to development of cloud computing, more and more local data are shared through cloud server by individuals/enterprises for specific purpose. However, in open networks and cloud environments that are not fully trusted, they face numerous security and privacy threats such as data leakage or disclosure, data corruption or loss, and privacy breaches when sending and receiving their data in a public cloud Recently, a number of studies have been conducted to address these risks, and a number of solutions have been proposed to enable data and privacy protection in untrusted cloud environments. This survey summarizes and analyzes state-of - the-art defense technologies in order to fully understand the developments and discover the research trends in this field. In this document, the security threats and specifications to protect the data services on the cloud server were presented and accompanied by a high-level description of the appropriate security techniques. This project, the symmetric encryption method was used to secure the data on cloud server. Peng at el. (2018) identified a detailed study of cloud database searchable encryption information. Information owners were inclined with the introduction of cloud storage to outsource their data to cloud providers. Regarding privacy concerns, prior to outsourcing, sensitive data should be encrypted. There are various encryption schemes that can be searched to ensure availability of data. However, for queries of data users, the existing search schemes are less efficient, especially for the multi-owner.

2.1 General Opinion about SE

Searchable encryption (SE) is an emerging research area in cloud computing. In this section, we review and analyze the existing searchable encryption schemes. SE can be divided into asymmetric searchable encryption and symmetric searchable encryption according to different cryptography primitives. In this paper, focus is on symmetric searchable encryption because usually it is computationally expensive to search for asymmetric encryption. To deal with symmetric searchable encryption, lots of researches have been done. Song et al. first defined the search problem on encrypted data and suggested a linear complexity symmetrical searchable encryption scheme. Goh et al. subsequently developed a safety concept for SSE and suggested a stable index based on pseudo-random functions and Bloom filters, but Goh's scheme's time cost is O(n). Curtmola et al. presented two formal SSE concepts and suggested a framework based on the inverted list to improve the performance of the request.

2.2 Research Gap

Author	Year	Title	Aim	Limitations
Ying et al.	2016	Adaptively secure cipher text-policy attribute-based encryption with dynamic policy updating	Securely search the data on cloud server using symmetric algorithm	Less secure Easy to access by authorized one's
Ding et al.	2017	Privacy-preserving multi-keyword top-k similarity search over encrypted data	Using symmetric searchable scheme which is based on pursue-random function to improve the security	More time used to complete the process and less Efficient and secure
Liu et al.	2017	Efficient symmetrical encryption to store dynamic social data from multiple sources on the cloud	To allow multiple parties to aggregate their data together while maintaining privacy	Limiting each participant's communication and computational complexity to a small constant
Peng et al.	2018	An easy multi-keyword search over encrypted cloud data for multiple data holders	Encrypted the sensitive data before outsourcing on cloud server using the symmetric algorithm	Less secure and More system disruption

3 EXISTING RESEARCH

Till date the problem has been related to searching of encrypted data on the cloud server and wished-for a symmetric algorithm for searchable encryption scheme. After that, Goh et al. came up with the searchable Symmetric encryption (SSE) and provided a safe keyword search based on the pseudo-random functions and the Bloom filters, but the time cost of the Goh scheme was O(n). Curtmola et al. then introduced SSE's two formal definitions and gave an inverted list-dependent method to improve query performance, which proved to be more efficient than the other works. Most of the work, however, scanned only the single keyword Boolean search support, which was not sufficiently advanced to support complex features. As a result, numerous works have been proposed in recent years to achieve various types of complex queries such as searching for similarities, searching multi-keywords, etc. The structure is shown in figure: 1.

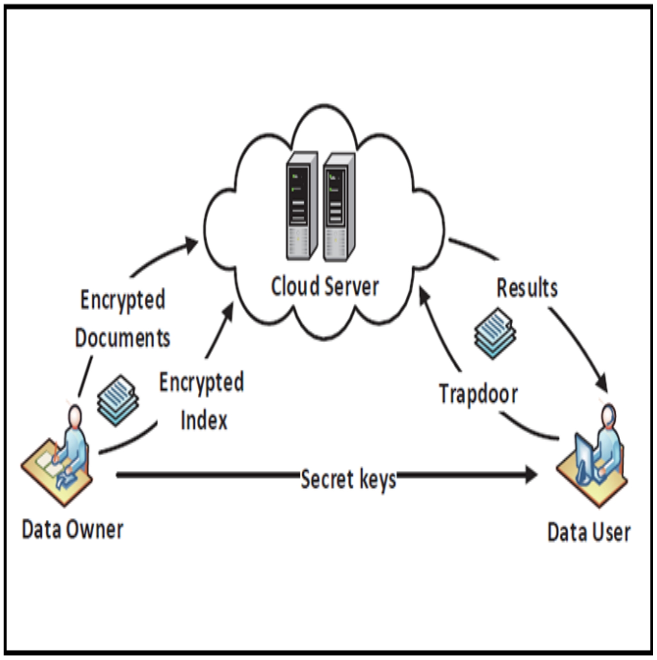

Figure 1. *Top-k search over encrypted data*

For the first time, Cao et al. proposed the multi-keyword search ranked over encrypted data and created a searchable index based on the vector space model, and selected "coordinate matching" to measure the similarity between queries and documents. Search time complexity is O(nm) where n is the number of keywords in the dictionary, m is the size of documents stored in the cloud server, and trapdoor construction time complexity is also very high.

3.1 Limitations of Existing System

Usually large costs in terms of data utility are caused by applying these approaches to data encryption, which makes traditional data processing methods work well over encrypted data. In this system, same key is used to secure the data; the one vulnerable to unauthorized access. The time complexity of creating trapdoor is high. Most of these approaches cannot simultaneously experience the high search efficiency and strong data protection, mostly when applied to large data authentication, which presents high scalability as well as the challenges of performance.

3.2 Existing System Mechanism

Symmetric cryptography: Symmetric cryptography is the cryptographic algorithm used for the purpose of encryption and decryption of the given data with an identical key. It is also referred to as secure key algorithm.

Data Encryption and Decryption: Encryption is the process of converting plain text, images or other information into a cipher text that is totally unreadable. The conversion of the encrypted data into the original form is called Decryption and is generally the reverse procedure of encryption. It helps to decode the encrypted information with a given secret key so that an authorized user can easily decrypt to read the data.

- Symmetric Algorithms: There are two types of symmetric algorithms:
- Block algorithms: in this system instead of encrypting data bit by bit, block of data is encrypted at one time. Length of a block varies depending upon the algorithm. System holds the data in the memory and waits for the complete blocks as the data is encrypted.
- Stream algorithms: in this system data is encrypted bit by bit and stored in the memory.
- Examples: DES, RC, AES Algorithm.

3.3 Flowchart of the existing mechanism

In existing system, there are three panel's admin, owner and user. The flowchart describes the working of existing system as shown in figure 2. The various steps are:

- User and owner register their detail and after that login on server.
- Admin login on server statically.
- Owner encrypts the selective file and uploads on server.
- Admin verify the file and gives the rank to that file.
- User searches the file on server via keyword and sends the request to admin for accessing purpose.
- Admin accept the request and send to owner for security key and owner send the security key to user
- . User uses that key to decrypt the data and to view and download the file.

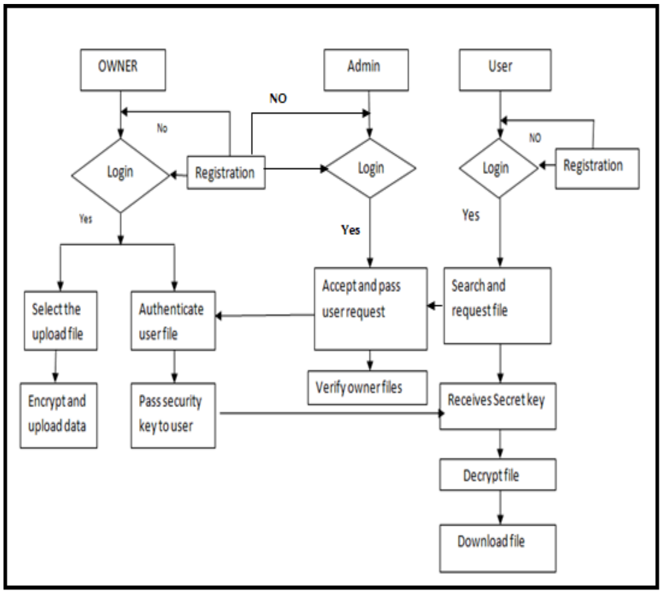

***Figure 2**. Flowchart of Existing System*

4 PROPOSED SYSTEM

The asymmetric encryption has been used to encrypt the data despite of symmetric encryption. The proposed system architecture is shown in figure 3.

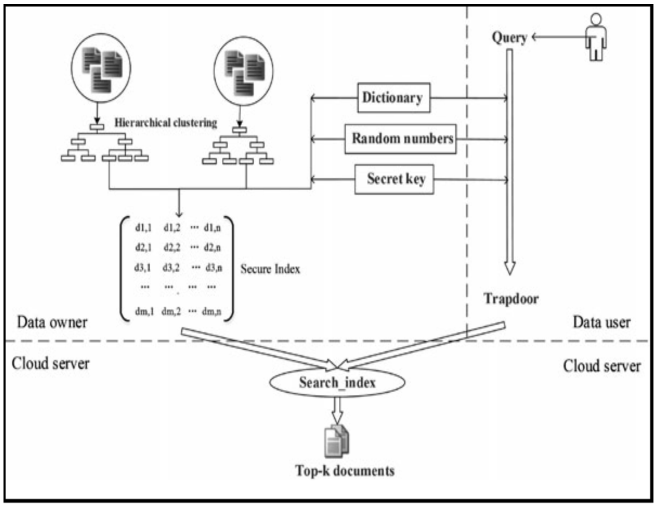

***Figure 3.** Proposed system architecture*

In addition, this depends on the hierarchical cluster and supports top-rank search over encrypted data to improve the query efficiency and data utility using the Group Multi-Keyword Top-k Search Scheme (GMTS). In this method, the cluster separates the dictionary sub-cluster and generates a searchable cluster list. For instance, using the Random Traversal Algorithmic Program (RTRA) in order to improve the data security, wherever the holder creates a hierarchical cluster as a searchable index and assigns a random key to the index, a random key is allocated to each file by the information client. In the proposed system, peer to peer architecture is used so as to improve the efficiency of communication between the user and owner.

4.1 Why Use Proposed System

First, we suggest the random traversal algorithm that would randomly traverse the cloud server on the index and give a different result for the same query, and it upholds the accuracy of unchanged queries for higher security. Depending on the random traversal algorithm, it represents a searchable encryption scheme that is both efficient and secure and can support top-k similarity search over the encrypted data. The data owner can control the query rate without sacrificing accuracy to unlink the capacity. The results show that the methods are more effective than the state-of - the-art methods and that data confidentiality can be better protected. Especially when dealing with the large data sets, the proposed method has good scalability efficiency.

4.2 Proposed System Mechanism

- Asymmetric cryptography

Asymmetric cryptography uses private and public keys to encrypt as well as decrypt data. Only key can be exchanged with all, so it's called the public key. The other key is kept as the private key is called a secret. Any key either public or private can be used for the encryption of the message. Message will be decrypted using the key other than that being used for encryption.

RSA asymmetric Algorithm

RSA algorithm is the asymmetric cryptography algorithm. The The data is encrypted and decrypted using the RSA algorithm with two different keys. i.e. Private Key and Public Key. As the name defines that the Public Key is given to everyone and Private Key is kept private.

4.3 Flowchart of the proposed mechanism

In the proposed system, there are three panels namely: admin, owner and user. The flowchart describes the working of the proposed system as shown in figure 4. The various steps are:

a) The user as well as the owner register their detail and then login to the server whereas admin login on server statically.
b) Admin verifies the users registered on server and send an activate key to user and the owner via mail.
c) Owner use the key provided to activate the account and encrypts the selective file and uploads on server.
d) There after the admin verifies the file and gives a rank to that file.
e) Similarly, the user also uses the key to activate the account and then search the file on the server using the keyword and rank to send the request to owner for private key.
f) Owner sends the security key to user.
g) User uses that key to decrypt the data and to view and download the file.

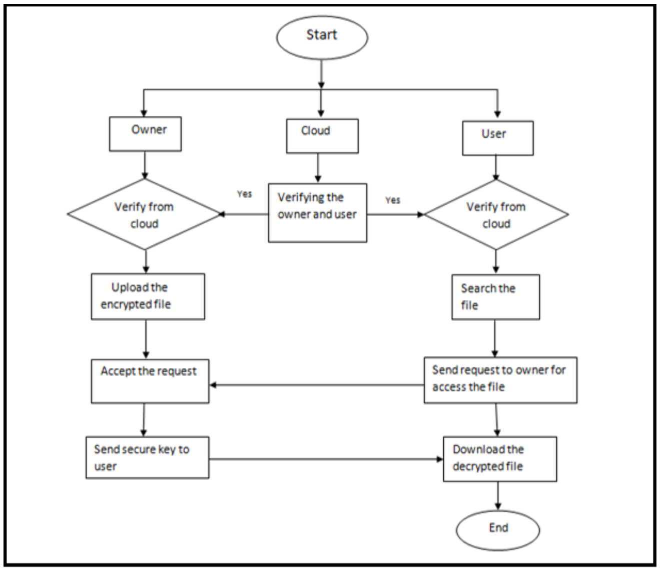
Start
Owner
Cloud
User
Verify from cloud
Yes
Verifying the owner and user
Yes
Verify from cloud
Upload the encrypted file
Search the file
Accept the request
Send request to owner for access the file
Send secure key to user
Download the decrypted file
End

***Figure 4**. Flowchart of proposed system*

4.4 Tools Used

Hardware Requirements:
System : Pentium Dual Core and above.
Hard Disk: 120 GB or more.
Monitor: LED
Input Devices: Mouse, Keyboard
Ram: 1GB

Software Requirements

Operating system: Windows7, 8.1,10.
Coding Language: HTML, CSS, PHP
Tool:NOTEPAD++
Platform web server: XAMPP
Database: MYSQL

5 CONCLUSION

Data security and privacy is the main concern in data storage for cloud computing. While cloud provides flexible and easily accessible data storage along with data management, there are still numerous possibilities for any malicious activity or interaction between intruders. Cloud server data may be confidential and therefore require more consideration of security. Cryptography techniques provide a safe and sound way for third-party confidential data storage using the encrypted singing form and provide only authorized users with the corresponding key. This paper gives an overview of the encrypted as well as the decrypted data in cloud server using the symmetric cryptography. It also gives an overview about the asymmetric cryptography. It helps the user to perform faster encryption and decryption of data without too much involvement of the cloud server. The anticipated system which we want to apply shows that the asymmetric search algorithm has been more secure than the symmetric encryption. It gives better security of the data from an unauthorized access. This application assures secure end to end search and transfer the data without any error.

REFERENCES

[1] Song, D. X., Wagner, D., & Perrig, A. (2000). Practical techniques for searches on encrypted data. In Proceeding 2000 IEEE Symposium on Security and Privacy. S&P 2000(pp. 44-55). IEEE.

[2] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-preserving multi-keyword ranked search over encrypted cloud data,” IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 1, pp. 222–233, 2014.

[3] Z. Xia, X. Wang, X. Sun, and Q. Wang, “A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data,” IEEE Transactions on Parallel and Distributed Systems, vol. 27, no. 2, pp. 340–352, 2016.

[4] C. D. Manning, P. Raghavan, H. Schutze ¨ et al., Introduction to information retrieval. Cambridge university press Cambridge, 2008, vol. 1, no. 1.

[5] S. Brin and L. Page, “The anatomy of a large-scale hypertextual web search engine,” Computer Networks and ISDN Systems, vol. 30, no. 17, 1998.

[6] Seth, S. M., & Mishra, R. (2011). Comparative analysis of encryption algorithms for data communication 1.

[7] Agrawal, M., & Mishra, P. (2012). A comparative survey on symmetric key encryption techniques. International Journal on Computer Science and Engineering, 4(5), 877.

[8] Fadhil Salman Abed, “A Proposed Method of Information hiding based on Hybrid Cryptography and Seganography”, International Journal of Application or Innovation in Engineering & Management, Vol. 2, Issue 4, April 2013

[9] Ajit Singh, Aarti Nandal and Swati Malik, “Implementation of Ceaser Cipher with Rail Fence for enhancing data Security”, International Journal of Advanced research in Computer Science and Software Engineering. Vol 2, Issue 12, December 2012 pp. 78 -82

[10] http://www.nku.edu (Fall 2006 Chris Christensen)

[11] Sinkov A., “Elementary Cryptanalysis – A Mathematical Approach”, Mathematical Association of America, 1996

[12] Vinod Saroha, SumanMor and Anurag Dagar, “Enhancing Security of Ceaser Cipher by Double Columnar Transposition

method", International Journal of Advanced Research in Computer Science and Software Engineering. Vol. 2, Issue 4, December 2012, pp. 39-49

[13] Kashish Goyal and Supriya Kinger "Modified Ceaser Cipher for Better Security Enhancement", International Journal of Computer Application, Vol. 73, Issue 3, July 2013, pp 26-31

[14] Cao, N., Wang, C., Li, M., Ren, K., & Lou, W. (2013). Privacy-preserving multi-keyword ranked search over encrypted cloud data. IEEE Transactions on parallel and distributed systems, 25(1), 222-233.

[15] Jung, T., Mao, X., Li, X. Y., Tang, S. J., Gong, W., & Zhang, L. (2013, April). Privacy-preserving data aggregation without secure channel: Multivariate polynomial evaluation. In 2013 Proceedings IEEE INFOCOM (pp. 2634-2642). IEEE.

[16] Yang, Y., Li, H., Liu, W., Yao, H., & Wen, M. (2014, December). Secure dynamic searchable symmetric encryption with constant document update cost. In 2014 IEEE Global Communications Conference (pp. 775-780). IEEE.

[17] Tayde, S., & Siledar, S. (2015). File Encryption, Decryption Using AES Algorithm in Android Phone. International Journel of Advanced Research in computer science and software engineering, 5(5).

[18] Saini, N., Pandey, N., & Singh, A. P. (2015, September). Enhancement of security using cryptographic techniques. In 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO)(Trends and Future Directions) (pp. 1-5). IEEE.

[19] K Srinivasa Rao , Dr Y. Vamsidhar (2015). Privacy-preserving multi-keyword ranked search over encrypted cloud data. International Journal of Applied Sciences, Engineering and Management ISSN 2320 – 3439(4),51 – 56.

[20] J. Baek, R. Safavi-Naini, and W. Susilo, "Public key encryption with keyword search revisited," in Computational Science and Its Applications. Springer, 2008, pp. 1249–1259.

[21] D. J. Park, K. Kim, and P. J. Lee, "Public key encryption with conjunctive field keyword search," in Information security applications. Springer, 2004, pp. 73–86.

[22] W. M. Liu, L. Wang, P. Cheng, K. Ren, S. Zhu, and M. Debbabi, "Pptp: Privacy-preserving traffic padding in web-based applications," IEEE Transactions on Dependable and Secure Computing, vol. 11, no. 6, Nov 2014

[23] Tang, J., Cui, Y., Li, Q., Ren, K., Liu, J., & Buyya, R. (2016). Ensuring security and privacy preservation for cloud data services. ACM Computing Surveys (CSUR), 49(1), 13.

[24] Ying, Z., Li, H., Ma, J., Zhang, J., & Cui, J. (2016). Adaptively secure ciphertext-policy attribute-based encryption with dynamic policy updating. Science China Information Sciences, 59(4), 042701.

[25] L. Arockiam, S. Monikandan "Data Security and Privacy in Cloud Storage using Hybrid Symmetirc Encryption Algorithm", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, Issue 8, August 2013, pp 3064-3070

[26] Mohit Marwahe, Rajeev Bedi, "Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing", International Journal of Computer Science Issues. Vol 10, Issue 1, January 2013, pp. 367-370

[27] Ding, X., Liu, P., & Jin, H. (2017). Privacy-Preserving Multi-Keyword Top-$ k $ k Similarity Search Over Encrypted Data. IEEE Transactions on Dependable and Secure Computing, 16(2), 344-357.

[28] Liu, C., Zhu, L., & Chen, J. (2017). Efficient searchable symmetric encryption for storing multiple source dynamic social data on cloud. Journal of Network and Computer Applications, 86, 3-14.

[29] Kisembe, P., & Jeberson, W. (2017). Future of Peer-To-Peer Technology with the rise of Cloud Computing. International Journal of Peer to Peer Networks (IJP2P), 8.

[30] Peng, T., Lin, Y., Yao, X., & Zhang, W. (2018). An efficient ranked multi-keyword search for multiple data owners over encrypted cloud data. IEEE Access, 6, 21924-21933.