iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** I **Month of publication:** January 2019

DOI: http://doi.org/10.22214/ijraset.2019.1132

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

Smart Search over Enciphered Data for Cloud Computing: A Survey

Sarvesh Kher[1], Mayuresh Patil[2], Gaurav Mahendrakar[3], Gauri Kavitkar[4], Nikhita Nerkar[5]
[1, 2, 3, 4]*Student, Dept. of Computer Engineering, RMD Sinhgad School of Engg, Pune, Maharashtra, India*
[6]*Asst. Professor, Dept. of Computer Engineering, RMD Sinhgad School of Engg, Pune, Maharashtra, India*

Abstract: ***With the advancement of information technologies particularly cloud storage used outsourcing data. Now a day's users store a large amount of data on the cloud but it's untrusted and we store secure data on the cloud. The concept of searchable encryption provides a promising direction in solving the privacy problem when*** *outsourcing data* ***to the cloud. Such schemes allow users to store their data in an encrypted format an untrusted server and then delegate the server to search on their behalf by issuing a trapdoor (i.e. encrypted keyword). This paper gives a detailed survey of various methods for searchable encryption schemes.***

Keywords: Encrypted Keyword, AES, Data Outsourcing, Error correction code rule, TF-IDF rule, SHA, Cloud Computing, Smart search over Enciphered Data.

I. INTRODUCTION

Now a day's users store an outsized quantity of information on the cloud however it's untrusted and that we store secure data on the cloud. With the advancement of data technologies, notably cloud storage services, info outsourcing and sharing became omnipresent in our life. as an example, a user Alice could store her information at Dropbox and share them together with her friends, within the meanwhile, she can also have access to her friends' information. Thanks to the non-public nature of private information, there's associate inherent would like for a user to by selection shares her information with completely different recipients. I observe, what a user will do is to line some access management policies so have confidence the cloud server (e.g. Dropbox) to enforce them. Sadly, this approach isn't realistic thanks to 2 reasons. One is that the users haven't any suggests that to stop the server from accessing their information. The opposite is that, albeit the server is benign, it should even be forced to share users' information with alternative parties (e.g. by the USA national Act). So it's needed to develop the thought of searchable encoding that provides a promising direction in finding the privacy drawback once outsourcing information to the cloud. Such schemes permit users to store their information in encoding from at Associate in nursing untrusted server so delegate the server to look on their behalf by supplying a personal key and encrypted search index.

II. LITERATURE REVIEW

In the work of Bao et al. [2], a brand new party, particularly user manager, is introduced into the system, to manage multiple users' search capabilities (e.g. alter them to look every other's data), during this extension, the user manager has to be absolutely sure since it's capable of submitting search queries and decrypting encrypted knowledge. This conflicts with our security criteria (i.e. there mustn't be further TTP involved).

In the work from [3], [4], and [5], the authors have investigated order-preserving encoding, wherever the ciphertexts preserve the order the plaintexts so each entity will perform Associate in Nursing equality comparison. Clearly, these schemes conjointly conflict with our security criteria (i.e. leak marginal data to the server).

Most existing rhombohedral searchable encoding schemes aim at permitting a user to source her encrypted knowledge to a cloud server and delegate the latter to go looking on her behalf. These schemes don't qualify as a secure and ascendable answer for the multiparty setting, wherever users source their encrypted knowledge to a cloud server and by selection authorize one another to go looking. Thanks to the chance that the cloud server might conspire with some malicious users, it's a challenge to own a secure and ascendable multiparty searchable encoding (MPSE) theme. this can be shown by our analysis on the Popa–Zeldovich theme, that says that Associate in Nursing honest user might leak all her search patterns although she shares only 1 of her documents with another malicious user. supported the analysis, the paper [6] presents a replacement security model for MPSE by considering the worst case and average-case situations, that capture totally different server-user collusion potentialities. Then they propose Associate in Nursing MPSE theme by using the additive property of Type-3 pairings and prove its security supported the additive Diffie–Hellman variant and rhombohedra external Diffie–Hellman assumptions within the random oracle model.

The [13] author introduces a system that develops associate economical looking formula associated with an economical matching formula within the CSP, so as to forestall, the man within the middle attack and impersonation attack. It secures key distribution formula from information the info the information} owner aspect to data user aspect, wherever the information coding keys is distributed in a very secure thanks to the information users.

The [14] paper proposes a completely unique secure search protocol permits that permits that enables completely different completely different knowledge house owners to code the files and indexes with different keys then construct a tree-based index structure for every knowledge owner and code with AOPPF and allows the cloud server to merge encrypted indexes while not knowing any info.

The [7] paper discuss the matter of privacy-preserving top-k keyword similarity search over outsourced cloud knowledge. Taking edit distance as a live of similarity, we tend to initial build up the similarity keyword sets for all the keywords within the knowledge assortment. We tend to then calculate the relevancy several the weather within the similarity keyword sets by the wide used TF-IDF theory. Leverage each the similarity keyword sets and therefore the relevancy scores, we tend to gift a brand new secure and economical tree-based index structure for privacy-preserving top-k keyword similarity search. To forestall potentially applied mathematics attacks, we tend to conjointly introduce a two-server model to separate the association between the index structure and therefore the knowledge assortment in cloud servers. Thorough analysis is given on the validity of search practicality and formal security proofs are bestowed for the privacy guarantee of our resolution. Experimental results on real-world knowledge set more demonstrate the provision and potency of our resolution.

The [12] author introduces a system during which stratified keyword search on remotely keep knowledge is completed by saving files in the cloud and retrieve the files by ransacking through the keywords. It's bestowed in stratified order victimization ranking formula within the index page. Security for knowledge keep in the cloud is completed through saving encrypted files and privacy of knowledge is maintained by providing completely different completely different trapdoors to different users. The stratified analysis is completed by score dynamics

The paper [8] propose a scientific resolution, that refers to as QDMiner, to mechanically mine question aspects by extracting and grouping frequent lists from free text, HTML tags, and repeat regions at intervals prime search results. Experimental results show that an outsized range of lists do exist and helpful question aspects may be deep-mined by QDMiner. They more analyze the matter of list duplication and find higher question aspects may be deep-mined by modeling fine-grained similarities between lists and penalizing the duplicated lists.

The author of [9] paper addresses issue by developing the fine-grained multi-keyword search schemes over encrypted cloud knowledge. They contribute the three-fold. First, they introduce the relevancy scores and preference factors upon keywords that modify the precise keyword search and customized user expertise. Second, they develop a sensible and really efficient multi-keyword search theme.

The planned theme will support difficult logic search the mixed “AND”, “OR” and “NO” operations of keywords. Third, they more use the classier sub-dictionaries technique to attain higher potency on index building, trapdoor generating and question. Lastly, they analyze the protection of the planned schemes in terms of confidentiality of documents, privacy protection of index and trapdoor, and UNLINK ABILITY of the trapdoor. Through in-depth experiments victimization the real-world dataset, they validate the performance of the planned schemes.

In [10] paper, the author considers objects that are labeled with keywords and are embedded in an exceeding vector area. For these datasets, they study queries that provoke the tightest teams of points satisfying a given set of keywords. It proposes a completely unique methodology known as ProMiSH (Projection and Multi-Scale Hashing) that uses random projection and hash-based index structures and achieves high measurability and acceleration. Also, they gift a definite associate degreed an approximate version of the formula. The experimental results on real and artificial datasets show that ProMiSH has up to sixty times of acceleration over progressive tree-based techniques.

The author of [11] paper proposes ECSED; a completely unique linguistics search theme supported the thought hierarchy and also the linguistic relationship between ideas within the encrypted datasets. ECSED uses 2 cloud servers. One is employed to store the outsourced knowledge sets and come to the hierarchic results to data users. The opposite one is employed to cipher the similarity scores between the documents and also the question and send the scores to the first server. To additional improve the search potency; they utilize a tree-based index structure to arrange all the document index vectors. Also, they use the multi-keyword hierarchic search over encrypted cloud knowledge because the basic frame to propose 2 secure schemes. The experiment results based on the $64000 world datasets show that the theme is additional EFFICIENT than previous schemes.

III. ALGORITHM USED

The summary of the various rule employed by the investigator within the previous paper is given below.

1) Encryption/Decryption victimization AES rule
2) Error correction code rule
3) TF-IDF rule
4) SHA-1

A. Encryption/Decryption victimization AES

1) AES relies on a style principle called a substitution-permutation network and is quick in each computer code and hardware. In contrast to its forerunner DES, AES doesn't use a Feistel network. AES could be a variant of Rijndael that includes a fastened block size of 128 bits, and a key size of 128 bits. in contrast, the Rijndael specification in and of itself is such as with block and key sizes which will be any multiple of thirty two bits, each with a minimum of 128 and a most of 256 bits.
2) AES operates on a 4×4 column-major order matrix of bytes termed the state, though some versions of Rijndael have a bigger block size and have further columns within the state. Most AES calculations are exhausted a special finite field.
3) The key size used for Associate in Nursing AES cipher specifies the amount of repetitions of transformation rounds that convert the input, known as the plaintext, into the ultimate output, known as the ciphertext. the amount of cycles of repetition is as follows:
4) 10 cycles of repetition for 128-bit keys.
5) Each spherical consists of many process steps, every containing four similar however completely different stages, as well as one that depends on the coding key itself. a collection of reverse rounds are applied to rework ciphertext into the initial plaintext victimization constant coding key.

B. ECC Algorithm

1) Elliptic-curve cryptography (ECC) is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. Elliptic curves can be used for encryption by combining the key agreement with a symmetric encryption scheme.
2) In today's world ECC algorithm is used in case of key exchanges by certificate authority (CA) to share the public key certificates with end users. Elliptic Curve Cryptography is a secure and more efficient encryption algorithm than RSA.
3) The security of ECC algorithm depends on its ability to compute a new point on the curve given the product points and encrypt this point as information to be exchanged between the end users.

The ECC system is based on the concepts of Elliptic Curves. To analyze the time taken by an algorithm researches have introduced polynomial time algorithms and exponential time algorithms. Algorithms with smaller computation can be evaluated with polynomial time algorithms and complex computations can be evaluated with exponential time algorithms.

The fig shows a simple elliptic curve.

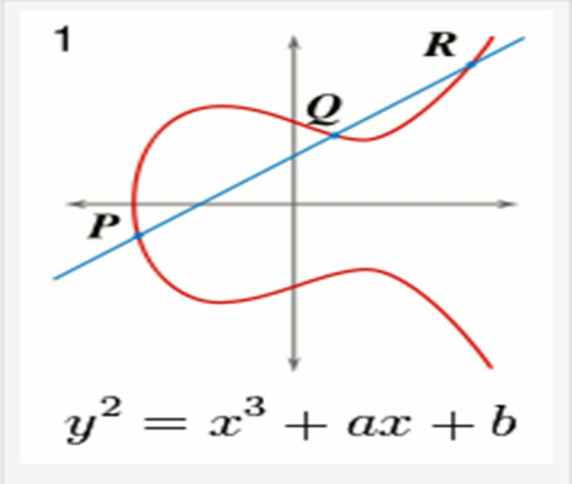


Figure 1: Simple Elliptic Curve.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue I, Jan 2019- Available at www.ijraset.com

The equation of an elliptic curve is given as,

$$y2 = x3 + ax + b$$

a) *Key Generation:* Key generation is an important part where an algorithm should generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key. Now, select a number, d within the range of n. Generate the public key using the following equation,

$$Q = d * P$$

Where d = the random number selected within the range of (1to n-1). P is the point on the curve, Q is the public key and d is the private key.

b) *Encryption*

Use the following equation to get back the original message 'm' that was sent.

M = C2 - d * C1

M is the original message that was sent

c) *Decryption*

Use the following equation to get back the original message 'm' that was sent.

M = C2 - d * C1

M IS THE ORIGINAL MESSAGE THAT WAS SENT

C. TF-IDF Algorithm

1) TF*IDF is an information retrieval technique that weighs a term's frequency (TF) and its inverse document frequency (IDF). Each word or term has its respective TF and IDF score. The product of the TF and IDF scores of a term is called the TF*IDF weight of that term.

2) The TF*IDF algorithm is used to weigh a keyword in any content and assign the importance to that keyword based on the number of times it appears in the document. More importantly, it checks how relevant the keyword is throughout the web, which is referred to as corpus.

For a term t in document d, the weight Wt,d of term t in document d is given by:

$$Wt,d = TFt,d \log (N/DFt)$$

Where:

a) TFt,d is the number of occurrences of t in document d.

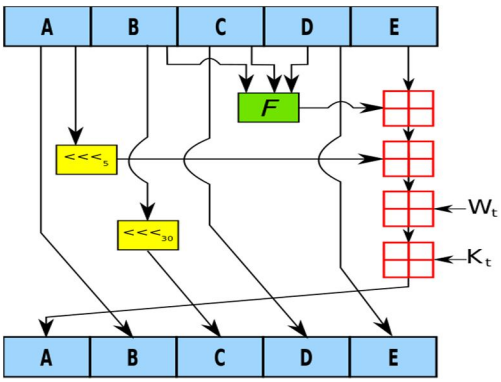
b) DFt is the number of documents containing the term t.

c) N is the total number of documents in the corpus.

D. Secure Hashing Algorithm-1

In cryptography, SHA-1 (Secure Hash Algorithm 1) is a cryptographic hash function designed by the United States National Security Agency and is a U.S. Federal Information Processing Standard published by the United States NIST. SHA-1 produces a 160-bit (20-byte) hash value known as a message digest. A SHA-1 hash value is typically rendered as a hexadecimal number, 40 digits long

1) *Image Description:* One iteration within the SHA-1 compression function: A, B, C, D and E are 32-bit words of the state; *F* is a nonlinear function that varies; *n* denotes a left bit rotation by *n* places; *n* varies for each operation; Wt is the expanded message word of round t; Kt is the round constant of round t; denotes addition modulo 232.

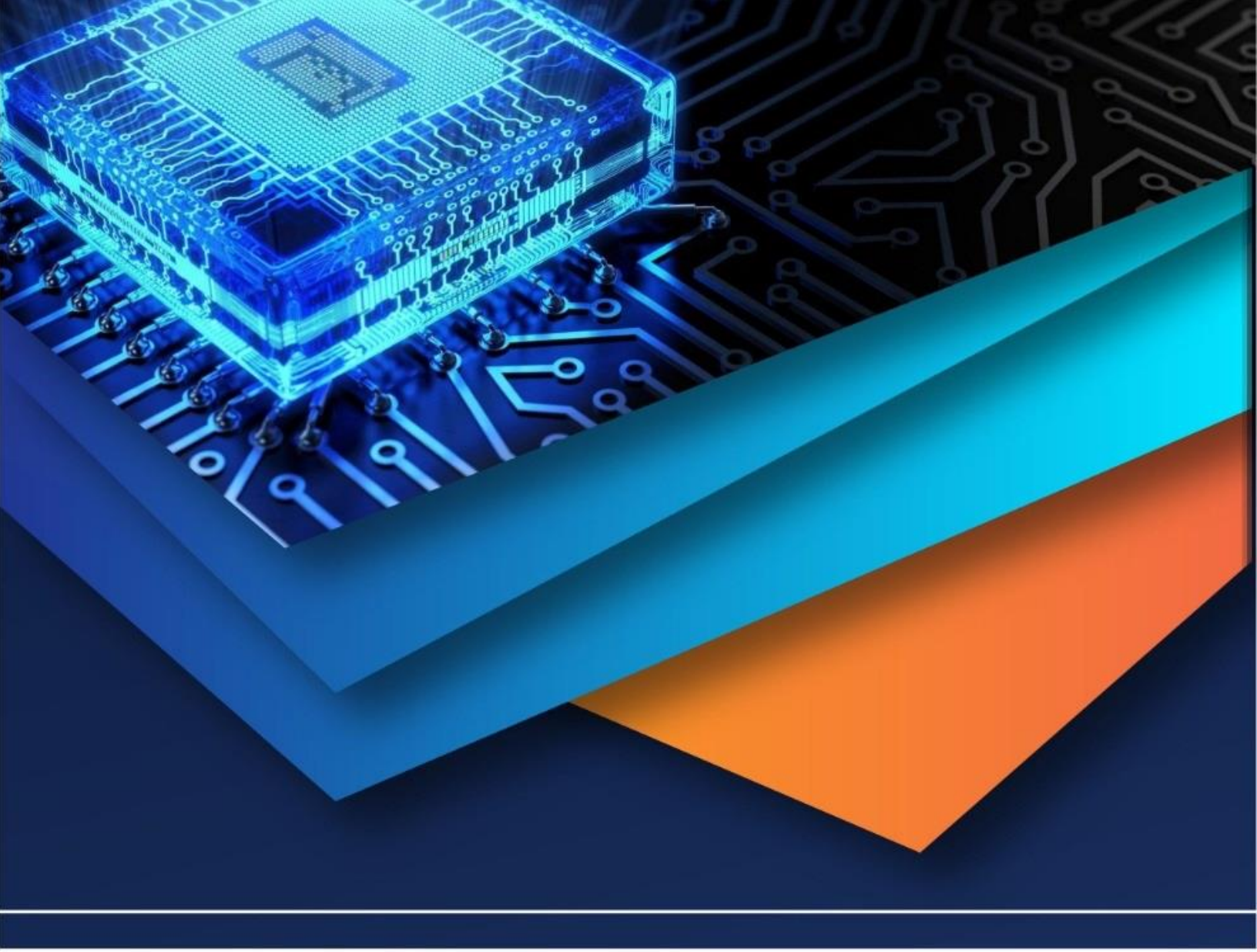


IV. CONCLUSION

To design a content-based search theme and build linguistics search more practical and context-aware could be a tough challenge. Several systems area unit projected to form encrypted knowledge searchable supported keywords. However, keyword-based search schemes ignore the linguistics illustration info of user's retrieval, and can't fully meet with users search intention. Here we tend to survey the various techniques won't looking out over encrypted knowledge. And that we return to understand that this system is simply able to search {the knowledge the info the information} over encrypted data however not in cloud computing. Therefore there's a desire to develop a system which may linguistics search {the knowledge the info the information} over encrypted data for cloud computing. Also, the system may be developed for knowledge storing and retrieving from the cloud with economical key management and sharing techniques

REFERENCES

[1] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: Improved definitions and efficient constructions," in Proc. 13th ACM Conf. Comput. Commun. Security, 2006, pp. 79–88.
[2] F. Bao, R. H. Deng, X. Ding, and Y. Yang, "Private query on encrypted data in multi-user settings," in Proc. 4th Int. Conf. Inf. Security Pract. Experience, vol. 4991. 2008, pp. 71–85.
[3] R.Agrawal,J.Kiernan,R.Srikant, and Y. Xu, "Order-preserving encryption for numeric data," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2004, pp. 563–574.
[4] A. Boldyreva, N. Chenette, Y. Lee, and A. O'Neill, "Order-preserving symmetric encryption," in Advances in Cryptology—EUROCRYPT (Lecture Notes in Computer Science), vol. 5479, A. Joux, Ed. Berlin, Germany: Springer-Verlag, 2009, pp. 224–241.
[5] D. X. Song, D. Wagner, and A. Perrig, "Practical techniques for searches on encrypted data," in Proc. IEEE Symp. Security Privacy, May 2000, pp. 44–55.
[6] Qiang Tang , "Nothing is for Free: Security in Searching Shared and Encrypted Data", IEEE Transactions On Information Forensics And Security, Vol. 9, NO. 11, November 2014.
[7] Teng Yiping, Cheng Xiang, Su Sen, Wang Yulong, Shuang Kai, "Privacy-Preserving Top-k Keyword Similarity Search over Outsourced Cloud Data", Dec 2015.
[8] Zhicheng Dou, Member, IEEE, Zhengbao Jiang, Sha Hu, Ji-Rong Wen, and Ruihua Song, "Automatically Mining Facets for Queries from Their Search Results", IEEE Transactions on Knowledge And Data Engineering, Vol. 28, No. 2, February 2016.
[9] Hongwei Li, Member, IEEE, Yi Yang, Tom H. Luan, Xiaohui Liang, Liang Zhou, and Xuemin (Sherman) Shen, "Enabling Fine-grained Multi-keyword Search Supporting Classified Sub-dictionaries over Encrypted Cloud Data", IEEE Transactions On Dependable And Secure Computing, Vol. 13, No. 3, May/June 2016.
[10] Vishwakarma Singh, Bo Zong, and Ambuj K. Singh, "Nearest Keyword Set Search in Multi-Dimensional Datasets", IEEE Transactions On Knowledge And Data Engineering, Vol. 28, No. 3, March 2016.
[11] Zhangjie Fu Lili Xia Xingming Sun Alex X. Liu Guowu Xie, "Semantic-aware Searching over Encrypted Data for Cloud Computing", IEEE Transactions on Information Forensics and Security, 2018.
[12] K Kiran Kumar, G Bharath Kumar, G Ramachandra Rao, Sk John Sydulu "Safe and High Secured Ranked Keyword Search over an Outsourced Cloud Data"in in International Journal of Research Volume 03 Issue 10 June 2017.
[13] Tianyue Peng, Yaping Lin, Xin Yao, Wei Zhang "An Efficient Ranked Multi-Keyword Search for Multiple Data Owners Over Encrypted Cloud Data" in Network Technology and Application
[14] N.Deepa, P.Vijayakumar, Bharat S. Rawal, B.Balamurugan "An extensive review and possible attack on the privacy preserving ranked multikeyword search for multiple data owners in cloud computing" in IEEE International Conference on Smart Cloud
[15] K Kiran Kumar, G Bharath Kumar, G Ramachandra Rao, Sk John Sydulu "Safe and High Secured Ranked Keyword Search over an Outsourced Cloud Data"in in International Journal of Research Volume 03 Issue 10 June 2017.
[16] Cong Wang, Ning Cao, Jin Li, Kui Ren, and Wenjing Lou "Secure Ranked Keyword Search over Encrypted Cloud Data" in 2010 International Conference on Distributed Computing Systems
[17] Ning Cao, Zhenyu Yang, Cong Wang, Kui Ren, and Wenjing Lou "Privacy-Preserving Query over Encrypted Graph- Structured Data in Cloud Computing" in 31st International Conference on Distributed Computing Systems
[18] Ayad Ibrahim, Hai Jin, Ali A. Yassin, Deqing Zou "Secure Rank-ordered Search of Multi-keyword Trapdoor over Encrypted Cloud Data" 2012 IEEE Asia-Pacific Services Computing Conference.
[19] Qiang Zhang, Yanhu Zhang, Jingyi Li "EasyComeEasyGo: Predicting bus arrival time with smart phone" 2015 Ninth International Conference on Frontier of Computer Science and Technology.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY