Dynamic Multi-Keyword Ranked Searchable Security Algorithm Using RC4+ and Forest

D, Palanivel Rajan Assistant. Professor, Department of Computer Science and Engineering, Coimbatore Institute of Engineering and Technology
palanivelrajan.d@gmail.com
Dr.S. John Alexis Professor Department of Automobile Engineering, Kumaraguru College of Technology

Abstract- Cloud gives plenty of recompense which makes enterprises to move their precious information to the cloud. The main reasons to movements are the simplified access methods, affordable cost which makes use of the storage service as important one. Since the user moves his/her valued data to the cloud so the security is the main input to create the conviction between the storage service provider and the user. To ascertain the security, the cryptography plays vital role by providing the searchable encryption. In survey presents the pros and cons of the various encryption techniques and different types of data structures. In this paper, we investigate multiple encryption methods for multi-keyword and propose the efficient searchable encryption schemes on the cloud platform.

Keywords: *Searchable Encryption, Multi keyword, cipher, Asymmetric encryption, tree, forest, Symmetric encryption, RC4+, minimum spanning tree.*

I. INTRODUCTION

Cloud computing is one of the best computing which shares the computing assets with numerous clients. The cloud has many advantage among that Information stockpiling is the one. This has the ability to extend towards the associations from the individual clients. [3]. The data was very much private on personal gadgets. In recent years a new term has evolved call "cloud" which is provided by different provides like Google Drive, iCloud, SkyDrive, Amazon S3, drop box and Microsoft Azure provide storage services., and which, platform, storage, software etc., and it is gaining importance because it frees the user from maintains perspective on an investment of some of money for the use of these services provided by cloud services.

Users tend to encrypt their data on the cloud using advanced encryption algorithms. In cloud computing, data owners may share their data in the cloud with authorized users who in turn might want to retrieve only the data files they are interested in [12]. Availability of required data at the right time and in the right format will be a key factor for gaining the acceptance of the end user. To retrieve a file over the cloud, keyword based retrieval is a trendy one in the recent day. Before this SSE schemes has deployed but this supports Boolean keywords search alone. That mean its checks either the keyword presents in the file or not. Later on the key word search was enhanced to include multiple-keywords.

In multi keyword based search index using top-k user ranking plays an important factor [10] . User ranking guarantee why something is

mentioned a lot. Search index is created for the files based on the user ranking. User ranking is an input to the user ranking. User ranking is an input to the cloud server and the retrieval of relevant files/data is performed by the cloud server depending on the ranking and the relevance score of the respective files [2].

In the remainder of this paper, the following information is presented: in Section III & IV, literature review in related area is discussed. Section V presents our proposed search schemes. Security analysis and performance analysis are presented in Section VI. Finally, in Section VII, the paper concludes with some suggestions for future work.

II. MOTIVATION

Still many security harms in cloud storage and which make the data and users feel insecure plenty of researches are going on this problem, towards this our contribution has summarized as below:

1. To provide an effective encrypted protocol for secure ranked keyword search over cloud data. [26], which fulfils the secure ranked search functionality with no relevance score information leakage against keyword privacy.
2. To make sure stable security, the asymmetric based ranked searchable encryption scheme with CRSA and B-tree has used. guarantee compared to previous searchable symmetric encryption (SSE) schemes.
3. Extensive experimental results demonstrate the effectiveness and efficiency of the proposed solution.

III. CRYPTOGRAPHIC

Cryptography comes as a branch in computer science and also origins from the mathematics. This deals with the data security and interrelated things along with more focus towards the encryption and authentication. In greek the word crypto means “hidden” while the word graphein mean “to write”. The plain text has converted into cipher in the encryption process. exactly the reverse operations take place in decryption process. i.e., the cipher text has converted into the plain-text.

Usually the encryption algorithms renovate data into unreadable form with the help of “KEY” and only those have the key only can decrypt the data. Generally, this is divided into two type symmetric and asymmetric key which is shown In Fig 1.

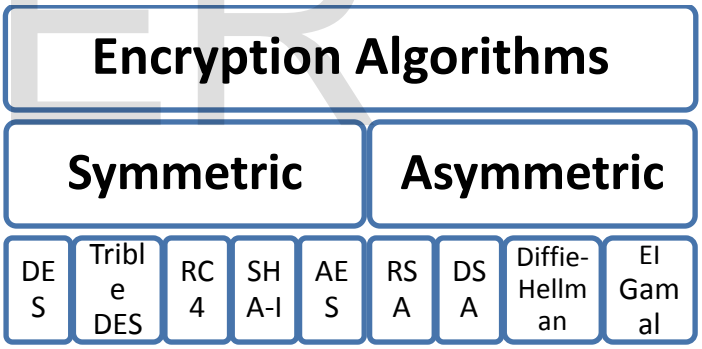


Figure 1 ENCRYPTION ALGORITHMS

Symmetric key encryption use only one key to encrypt and decrypt the data different symmetric algorithms is discussed in the Table I Different symmetric encryption algorithms. The Structure, Key Size , Rounds, Cipher Type are listed in the Table II. In another hand asymmetric key encryption uses two keys i.e., private and public keys are used.

Public key is used for encryption and private key is used for decryption which is

TABLE I Different Symmetric encryption algorithms.

AES	*SHA-1*	*DES*	*Triple DES*	*RC4*
The Advanced Encryptions Standard (AES) is a symmetric key encryption/decryption algorithm for converting plain-text to cipher text and vice-versa [24]. Since the same key or master key is used, the must be kept secret or with trusted 3rd party, because compromise of this key would mean compromise to the data.	SHA stands for "Secure Hash Algorithm", SHA-1 is cryptographic hash function technique where hash of data is computed [6]. AS compared to SHA-0, SHA-1 is widely used because it corrects errors in SHA hash specification, which led to weakness.	The Data Encryption Standard (DES) has been the one of strongest symmetric-key algorithm for data encryption. It has high influence in the recent cryptography domain. [17].	Due to the increasing computational capacity the original 56 bit DES cipher key can be cracked using the brute-force attacks. So to avoid these issues the Triple DES methods increase the key size without redesigning the block cipher algorithm. [14].	Stream cipher based Rivest Cipher 4 (RC4) has used in the Transport Layer Security (TLS). RC4 could be vulnerable while starting of the output key stream has not discarded and some ways of using RC4 could lead to very insecure protocols such as WEP [8].

discussed in the Table III Different Asymmetric encryption algorithms.

A block cipher is a deterministic and computable function of k-bit keys and n-bit (plaintext) blocks to n-bit (cipher text) blocks.

TABLE II Symmetric encryption algorithms Parameters

Algorithm	**Structure**	**Key Size (In bits)**	**Rounds**	**Cipher Type**
AES	Substitution-permutation network	128,192,256	10,12,14	Block
DES	Balanced Feistel network	56	16	Block
Triple DES	Feistel network	112,168	48	Block
SHA-I	Merkle–Damgård construction	160	80	Block
RC4	-	40 to 2064	1	Stream

when we encrypt the same plaintext block with the same key, we'll get the same result.

A stream cipher has a function which directly maps k-bit keys and arbitrary length plaintexts to the cipher ext. The prefixes of the plaintext map to prefixes of the cipher text. Due to this the starting part of the cipher text can be computed before the trailing part of the plaintext is known.

The blocks don't have to be bit-sized, n-character-blocks would fit here. This means,

IV. DATA STRUCTURES

A connected graph with no cycles is called as tree and a graph with each connected components of tree is called as forest, which is shown in the figure 2. A leaf in a tree is any vertex of degree 1. Consider any leaf of T. This vertex is adjacent to exactly one edge. Remove

this vertex and edge contributing 1 each to the number of vertices and edges. Continue removing leaf / edge pairs until we are left with just a single edge. A graph with a single edge has one more vertex than edge, hence the total number of edges is one less than the total number of vertices.

TABLE III Different Asymmetric encryption algorithms

RSA	DSA	Diffie-Hellman Key Exchange (D-H)	EI Gamel
This is a web secret authentication system that uses an algorithmic program developed in 1977 by Ron Rivest, Adi Shamir, and author Adleman. The RSA algorithmic program is that the most typically used in secret writing. until currently it's the sole algorithmic program used for personal and public key generation and secret writing. it's a quick encrypion [32].	The Digital Signature algorithm (DSA) may be a Federal science normal for digital signatures. it absolutely was projected by the National Institute of normals and Technology (NIST) in August 1991 to be used in their Digital Signature Standard (DSS) With DSA, the entropy, secrecy, and individualism of the random signature price k is crucial [31]. it's therefore crucial that violating anyone of these 3 necessities will reveal the complete personal key to an assaulter. Exploitation a similar price doubly (even whereas keeping k secret), employing a inevitable value, or leaky even many bits of k in every of many signatures, is enough to interrupt DSA.	Diffie–Hellman key exchange may be a specific methodology of exchanging cryptologic keys. it's one among the earliest sensible samples of key exchange enforced inside the sector of cryptography. The Diffie–Hellman key exchange methodology permits 2 parties that haven't any previous data of every different to collectively establish a shared secret key over an insecure communications channel. This key will then be accustomed inscribe consequent communications employing isobilateral key cipher.	In cryptography, the ElGamal secret writing system is an uneven key secret writing algorithmic program for public-key cryptography that relies on the Diffie–Hellman key exchange. it absolutely was represented by Taher Elgamal in 1984. ElGamal secret writing is employed within the free antelope Privacy Guard package, recent versions of PGP, and different cryptosystems. The Digital Signature algorithmic program may be a variant of the ElGamal signature theme, that mustn't be confused with ElGamal secret writing. ElGamal secret writing will be outlined over any cyclic cluster . Its security depends upon the issue of a precise problem in associated with computing distinct algorithms.

If there exists of the graph G has planar and that is embedding of G into the plane, then no two edges will clash each other [1]. For any tree T = (V, E) with |V | = n, |E| = n – 1

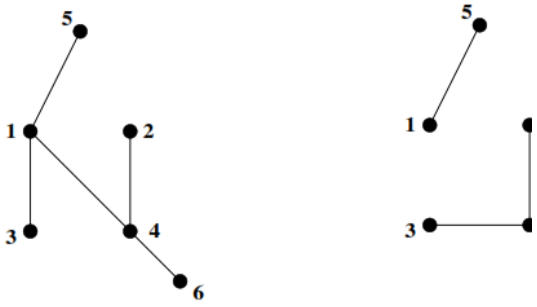


Figure 2 A TREE AND A FOREST

A. Binary tree

When each node has at most two children's then that tree is called as binary tree. These two children are called as left child and right child [18]. A recursive definition using just set theory notions has that a binary tree in a triple (L, S, R), where L and R have the binary trees or the empty set and S has the singleton set.[4] Some cases the binary tree considered to be the empty set as well.

B. Binary search tree

Sorted binary tree also called as Binary search tree (BST), it's a particular type of storage in these Abstract structures which contains the elements like name, number, etc., in the memory.

This enables the fast lookup, insertion and deletion of the items. This can be implemented using two different way one is lookup tables and another one is dynamic sets of items. This allow the searching a elements by using its key. For example, the telephone number of the person can be find using his/her name. [13].

C. AVL tree

Georgy **A**delson-**V**elsky and **L**andis' tree called as the AVL tree. It's the first kind of binary search tree data structure, which has the ability of self-balancing. The height of the left and right of the sub tree of any node differ by at most one. If at any time they differ by more than one, rebalancing is done to restore this property. The time taken for the search, adding and removing of the elements for the average and worst case is O(log n) . Where N is the number of nodes in the tree before the operation performed. Rebalancing has required during the addition and removal of the nodes in tree. This can be done by rotating one or more nodes in the tree. [16].

D. Minimum span tree

In a spanning tree has all the vertices covered with minimum possible number of edges then that's called as minimum spanning tree. It doesn't have the cycles and it's not possible to disconnect. Each edge is assigned with the weight. This assigned weight of each edge is computed to get the sum of the weights of that spanning tree. [7].

E. Forest

Forest is a set of ordered general trees. The tree is the root of the forest. In this nodes might have two or more children's [11]. the trees of the forest that it roots i.e., The children of a node have sequenced as first, second and etc. Notice that a general tree must have a root (in contrast to a binary tree), and that a forest may be empty (it is a set, and sets can be empty).

V. PROPOSED SEARCH SCHEME

For our system, we choose the Minimum spanning tree as indexing data structure to identify the match between search query and data documents. Specially, we use inner data correspondence, i.e., the total number of query keywords appearing in file to evaluate the similarity of that file with the given search query. Each keyword are converted to the tree and these trees forms the corresponding document as a forest and the whole indexed has encrypted using $RC4^{+}$ cipher (RCPF). Whenever user wants to search, he/she creates a trapdoor for the keywords.

We have designed and analyzed the performance of multiple keywords ranked search scheme using $RC4^{+}$ algorithm and forest data structure for searchable index tree. Further, we analyzed its performance over similarity based multiple keywords search (SBMKS). We have used CloudSim platform to simulate the proposed system and to study its performance.

RC4 modified with more three-phase key scheduling is called as RC4+. This is taking about 3 times more than RC4 for the output function. In RC4+ also performs four additional lookups in the S array for each byte output, which takes 1.7 times more than the RC4.

```
void BuildTree(Document G, Keyword mst[])
 {
int i, k, v, w;
Edge a[MAXE];  // list of all words in G
int E = getkeywords(a, G); // keywords in G
sort(a, 0, E-1); // sort keywords by weight
 UFinit(G->V);
 for (i = k = 0; i < E && k < G->V-1; i++)
{
 v = a[i].v;
 w = a[i].w;
 // if keyword a[i] doesn't create a cycle, add to tree
 if (!find(v, w)) {
union(v, w);
mst[k++] = a[i];
}
 }
}
Void BuildForest(Document G){
Tree t[],cnt=0;
For (i=0:i<G;i++) // no of possible keywords
{
        While(key!=0)
           //Split the keywords as parts
           t[cnt]= BuildTree(G[i], key);
}
}
```

```
Void rc4p(Forest S){
while GeneratingKey {
   i := i + 1
   a := S[i]
   j := j + a
   b := S[j]
   //(Swap S[i] and S[j])
   S[i] := b
   S[j] := a
   c := S[i<<5⊕ j>>3] + S[j<<5 ⊕ i>>3]
   key (S[a+b] + S[c⊕0xAA]) ⊕ S[j+b]
}
}

Void main(){
Forest F[];
 For(j=0;j<k;j++){ // k is no of the documents
          F[j]=BuildForest();
          Rcp4p(F[j]);
          UpdateIndex();

}
}
```

Usually normal RC4+ uses modulo 256 but our proposed methods use the 512 bit one to ensure the fast operations.

TABLE VI Proposed System Model

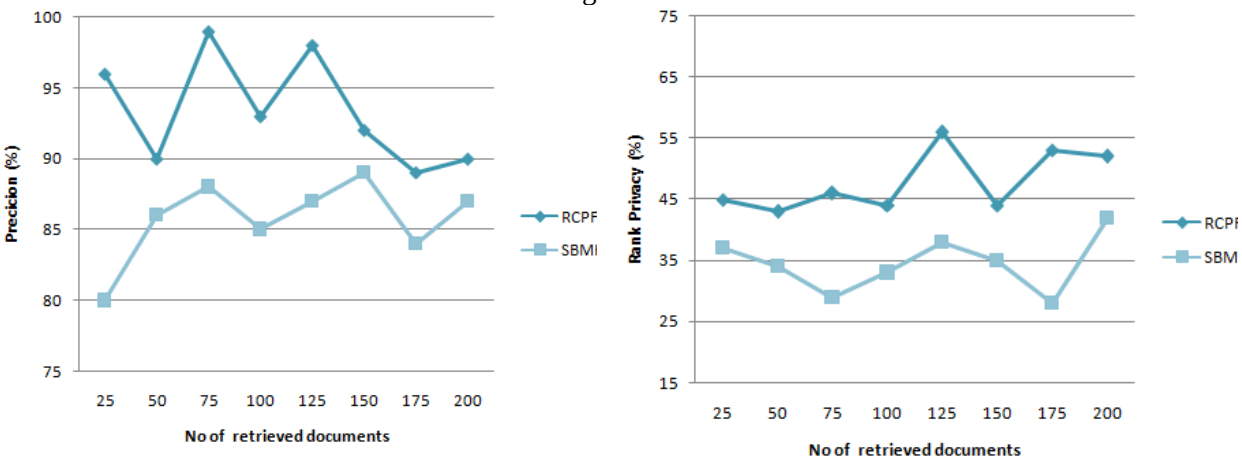
VI. PERFORMANCE ANALYSIS

The security of the designed system is provided by using RC4+. As long as private key (encrypted) is kept secret the cloud provider cannot deduce index tree or documents set. Since trapdoor is also encrypted using RC4+, the provider cannot make out the keywords inside the trapdoor maintaining the confidentiality at index and query level. The documents in cloud storage are also protected, since documents are encrypted using RC4+. Without having the decryption key, it is highly hard to decrypt the documents thus provides security at storage level.

To be useful and usable, databases must support operations, such as search, deletion and insertion of data. For large organizations the databases are huge in size and cannot be maintained entirely in memory. By using spanning forest to construct the index for the data we can improve the search efficiency. Forest minimizes the disk I/O (disk read and disk write) by copying a block of data (page) containing many records at a time into memory. This in turn improves the search efficiency. Asymptotically, searching an unsorted database without indexing will have a worst case running time of O(n), where n represents the number of keywords. If the same data is indexed with a Forest, the same search operation will run in logarithmic time i.e. O (log n).

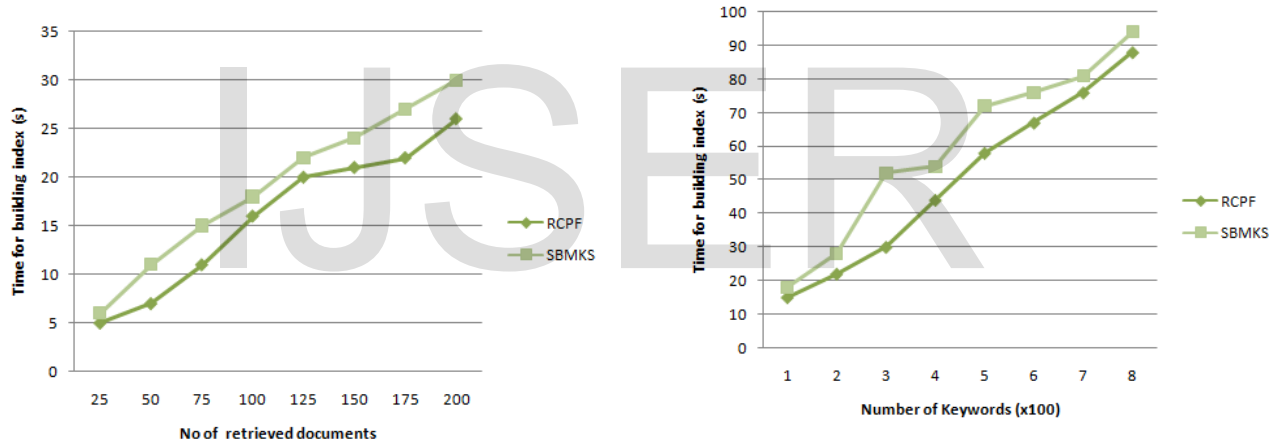
Figure 3 depicts the Precision and rank privacy. our proposed system RCPF with SBMKS. In this study, we compared the performance of Results clearly show that our proposed RC4+

Figure 3



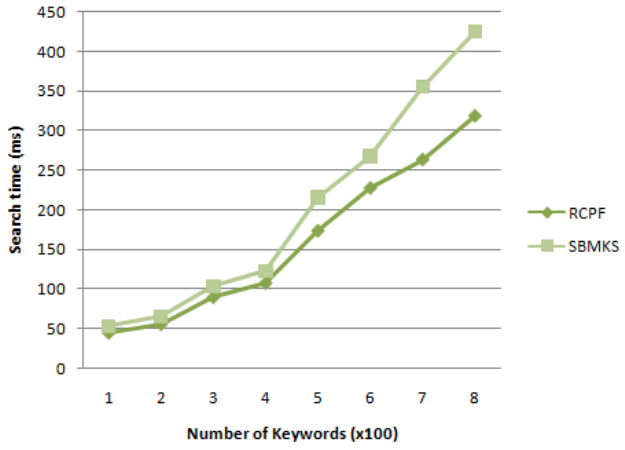
(a) Precision (b) Rank privacy

Figure 4

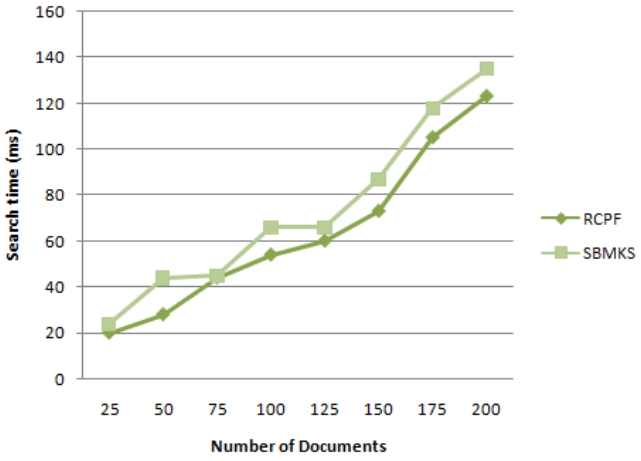


(a) Indexing Time for different Documents (b) Indexing Time for different number of keywords

Figure 5



(a) Search Time for different Documents



(b) Search Time for different number of keywords

based scheme performs better even under increase number of documents.

In the Figure 5 plotted makes the comparison of the search time in milliseconds of our proposed system RCPF against the SBMKS. For searching, the time taken by the SBMKS is approximately 54 milliseconds, whereas our proposed system takes approximately 45 seconds. As the number of keywords increased for searching, the searching time also increases in both system, however when compare to the time the RCPF scheme is found to be better. Thus it is evident that encryption algorithm RC4+ with Forest as index tree performs better than SBMKS.

VII. Conclusion

This work uses RC4$^+$ algorithm for encrypting data files and index tree based on Forest. RC4$^+$ increases the data security and improves privacy of data by its commutative nature. Using RC4$^+$, data in a file can be updated dynamically without affecting the overall performance of searching on B-tree. In our proposed system, if encrypted data is modified, re-encrypting for the whole data is not needed. This is a desirable feature as it reduces the computation time and RCPF increase the precision of the search also. This reduce the indexing time and build the index fast when compare to the SBMKS and also give the enhanced performance in the search even the keywords or documents increase also.

The future work would concentrate on using graph theory and Elliptic Curve Diffie-Hellman (ECDH) encryption technique for better performance. Further, we intend to analyze the behavior of our proposed system(s) for multiuser environment.

References

[1] C.J.Stam, P.Tewarie, *"The tree and the forest: Characterization of complex brain networks with minimum spanning trees", in Elsevier International Journal of Psychophysiology 92 (2014) 129-13.*

[2] Ranjeet Masram, Vivek Shahare, *" Analysis and Comparison of Symmetric Key Cryptographic Algorithms Based on Various File Feature", in International Jouranl of Network Security & Its applications, Vol.6, No.4, July 2014*

[3] B.J.Jonkheid, R.A.Roebeling, *"A fast SEVIRI simulator for auantifying retrieval uncertainties in the CM SAF cloud physical property algorithm", in Atmospheric Chemistry and Physics, 2012*

[4] Cong Wang, Kui Ren, *"Privacy-Preserving public auditing for Secure cloud storage", In IEEE Transactions on computers, Vol.62, No.2, 2013*

[5] *C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-Preserving Public Auditing for Storage Security in Cloud Computing," Proc. IEEE INFOCOM '10, Mar. 2010.*

[6] *P. Mell and T. Grance, "Draft NIST Working Definition of Cloud Computing," http://csrc.nist.gov/groups/SNS/cloudcomputing/index. html,June 2009.*

[7] *M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.H. Katz, A.Konwinski, G. Lee, D.A. Patterson, A. Rabkin, I. Stoica, and M.Zaharia, "Above the Clouds: A Berkeley View of Cloud Computing,"Technical Report UCB-EECS-2009-28, Univ. of California, Berkeley, Feb. 2009.*

[8] *Cloud Security Alliance, "Top Threats to Cloud Computing," http://www.cloudsecurityalliance.org, 2010.*

[9] *Wenhai Sun, Ning Cao, "Verifiable Privacy-Preserving Multi-Keyword Text Search in the Cloud Supporting Similarity-Based Ranking", In IEEE Transactions On Parallel and distributed System, Vol 25, No.11, November 2014.*

[10] *S. Zerr, D. Olmedilla, W. Nejdl, and W. Siberski, ''Zerber+r: Top-k Retrieval from a Confidential Index,'' in Proc. EDBT, 2009, pp. 439-449.*

[11] *Cloud Security AllianceSecurity Guidance for Critical Areas of Focus in Cloud Computing V3.0. [Online]. Available: http://www.cloudsecurityalliance.org 2011.*

[12] *Zhangjie Fu, Xingming Sun,"Achieving Effective Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Synonym Query ", in IEEE Transactions on Consumer Electronics, Vol. 60, No. 1, February 2014*

[13] *P.A. Cabarcos, F.A. Mendoza, R.S. Guerrero, A.M. Lopez, and D. DiazSanchez,"SuSSo: seamless and ubiquitous single sign-on for cloud service continuity across devices,"IEEE Trans. Consumer Electron.,vol.58, no. 4,pp.1425-1433,2012.*

[14] *D. Diaz-Sanchez, F. Almenarez, A. Marin, D. Proserpio, and P.A. Cabarcos,"Media cloud: an open cloud computing middleware for content management," IEEE Trans. Consumer Electron., vol. 57, no. 2, pp. 970-978, 2011.*

[15] *S. G. Lee, D. Lee, and S. Lee, "Personalized DTV program recommendation system under a cloud computing environment," IEEE Trans. Consumer Electron., vol. 56, no. 2, pp. 1034-1042, 2010.*

[16] Weerasinghe, T.D.B, "An effective RC4 stream cipher", 2013 8th IEEE International Conference on 2013

[17] J. Xie, X. Pan," An Improved RC4 Stream Cipher", 2010 International Conference on Computer Application and System Modeling, (ICCASM 2010), pp. (V7) 156-159, 2010 .

[18] S. S. Gupta, A. Chattopadhyay, K. Sinha, S. Maitra, B. Sinha," High Performance Hardware Implementation for RC4 Stream Cipher", Computers, IEEE Transactions on, vol. 62, no. 4, pp. 730,743, April 2013 doi: 10. 1109/TC. 2012.

[19] S. Fluhrer, I. Mantin, A. Shamir, "Weaknesses in the Key Scheduling Algorithm of RC4", SAC2001 (S. Vaudenay, A. Youssef, eds.), vol. 2259, no., pp. 1,24, Springer, Verlag, 2001

[20] Das.S,Dey, H. Ghosh, R,Computer, "An approach to assess the optimality of refining RC4", Communication, Control and Information Technology (C3IT), 2015 Third International Conference on 2015.

[21] Jian Xie, Xiaozhong Pan, "An improved RC4 stream cipher", Computer Application and System Modeling (ICCASM), International Conference on 2010.

[22] Qian Yu,Zhang, Orumiehchiha, Hua Li, "RC4-BHF: An Improved RC4-Based Hash Function", in Computer and Information Technology (CIT), 2012 IEEE 12th International Conference on 2012

[23] Jindal,Singh, "Performance analysis of modified RC4 encryption algorithm", on Recent Advances and Innovations in Engineering (ICRAIE), 2014

[24] Yanling Xing ,Yukui Pei, Ning Ge, "LT code design based on RC4 sequential cipher", on Communication Technology (ICCT) IEEE 14th International Conference on 2012

[25] Srinivas, S, Biswas, N.N. Massively , "A fast algorithm for data exchange in reconfigurable tree structures", Proceedings of the First International Conference on 1994

[26] Dalbouh, H.A., Norwawi, N.M, "Improvement on Agglomerative Hierarchical Clustering Algorithm Based on Tree Data Structure with Bidirectional Approach", on Intelligent Systems, Modelling and Simulation (ISMS), Third International Conference on 2012

[27] Bucciarelli, A, Salibra, A on Logic in Computer Science, "The sensible graph theories of lambda calculus", Proceedings of the 19th Annual IEEE Symposium on 2004.

[28] Delbem, de Lima, T.W, Telles, G.P on Evolutionary Computation, "Efficient Forest Data Structure for Evolutionary Algorithms Applied to Network Design", IEEE Transactions on 2012

[29] P. M. S. Carvalho , L. A. F. M. Ferreira and L. M. F. Barruncho "On spanning-tree recombination in evolutionary large-scale networkproblems: Application

to electrical distribution planning", IEEE Trans. Evol.Computat., vol. 5, no. 6, pp.623 -630 2001 .

[30] H. H. Chou , G. Premkumar and C. H. Chu "Geneticalgorithms for communications network design: An empirical study of the factors that influenceperformance", IEEE Trans. Evol. Computat., vol. 5, no. 3, pp.236 -249 2001

[31] Lane, R.O, Cooper, T.M. ; Maskell, S.R, "Efficient data structures for large scale tracking", on Information Fusion (FUSION), 17th International Conference on 2014.

[32] Gurpreet Singh, Supriya Kinger"Integrating AES, DES, and 3-DES Encryption Algorithms for Enhanced Data Security "International Journal of Scientific & Engineering Research, Volume 4, Issue 7, July-2013.

[33] Uma Somani, "Implementing Digital Signature with RSA Encryption Algorithm to Enhance the Data Security of Cloud in Cloud Computing," 2010 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010).