Large Scale Data Storage and Retrieval System using Keywords for E-governance

Dr. Bondu Venkateswarlu[1], Soumya G.V[2]
[1]Department of Computer Science and Engineering,
Dyanda Sagar University,
Bangalore, India.
[1]*iambondu@gmail.com*
[2]M.Tech, Department of Computer Science and Engineering,
Dyanda Sagar University,
Bangalore, India.
[2]*gvsoumya9@gmail.com*

***Abstract*—**whenever we are storing any documents related E governance to cloud we need to secure it because there may be a malicious or third party attackers can hack data which in stored in cloud so that we need to create dynamic secure storage system. We are proposing improved encryption algorithm to give security for the text files related to E governance which we are going to store in the cloud. The improved RNS algorithm is one of the most widespread and commonly used methods and intriguing in the distribution and exchange of key. It can be said that this algorithm is a public key encryption system; the only objective of this algorithm is key distribution. After achieving security, to retrieve the file from the server Speed of transmission should be high, to achieve speech of transmission and searching efficiency, we are proposing Index Array Structure using MD5 Algorithm and Multi keyword Search for Searching Efficiency. To achieve Index Array Structure, from the text file all the unnecessary words, special characters and whitespaces are removed. The remaining keywords are extracted from the file. Remaining keywords occurrence is checked how many times the keywords are repeated. The repeated keywords weight age is noted in index array table. After weightage calculation, keyword ranking is checked. Keyword ranking is nothing but number of occurrences of keyword divided by total number of keywords in particular files (Term frequency). Finally, keyword ranking is stored in index array. After completing this process all the keywords are converted into hashed index with the grade access control key using MD5 algorithm. This converted hash key is inserted into index array. Finally, with the help index array content we are going to achieve Searching efficiency.

Key words*-Message Digest (MD5), Improved Residue Number System (RNS), hash key, index array*

*****

I. INTRODUCTION

These days, Government applications are used to entrance the information with the rebellions changes and make the people to process through internet in a meek and well-organized way, in terms of interrelate and study. Naturally variations of the government functions imitate in the government organization, relationship between people and government, big business and organizations etc. In cloud computing, E-Governance leads a major role of client appropriate and cost savings. The infrastructure of software and hardware involves the usage over the Internet for hosting the applications distantly. E-Governance application makes easy to do the actions like income taxes, pension services, administration etc. by means of IT infrastructure. It improves the idleness reduction at various levels with the function efficiency. It provides convenience of various services in the framework of E-Governance regardless of language barricades and the locations. In existing models various problems of framework is defined and incapable all categories to address from countryside to metropolitan citizens.

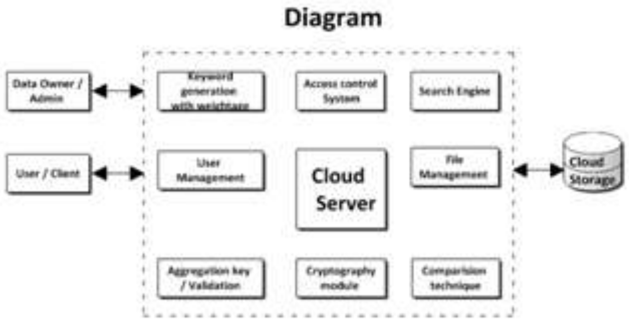

Figure 1: E-governance application modules

The above figure it shows the overall architecture of the E-governance application where it consists of two modules user and client. Here the user can access the files, the module it consists of key word generation, user management, validation, file management, cloud server and cloud storage.

II. RELATED WORK

Cloud storage allows users to distantly store and recover their data and enjoy the on-demand high quality cloud applications without the burden of local hardware and software management. [1]Cloud storage system allows storing of data in the cloud server efficiently and makes the user to work with the data without any trouble of the resources. Cloud computing is highly proficient technology because of its unlimited resource provisioning and data storage services which help us in managing the data as per necessities. In the current situation in India we are keen to implement the e-Governance model. The town areas are in good position in form to avail the services of e-Governance as they have all the obligatory infrastructure but in rural areas the biggest problem is the non-availability of proper organization as well absence of computer aware citizen, Cloud Computing can be a upcoming solution to achieve that needs[2].Using Cloud Storage, users can distantly store their data and enjoy the on-demand high quality applications and facilities from a shared pool of configurable computing resources, without the load of local data storage and preservation. Though, the fact that users no longer have physical possession of the outsourced data makes the data integrity protection in Cloud Computing a difficult task, especially for users with inhibited computing resources. [3]Moreover, users must be able to just use the cloud storage as if it is local, without worrying about the need to verify its truthfulness. A cloud storage organization, containing of a collection of storage servers, provides long-term storing services over the Internet. Storing data in a third party’s cloud system causes serious concern over data discretion.[4] General encryption schemes protect data privacy, but also limit the functionality of the storage system because a few processes are supported over encrypted data. Building a secure storage system that supports multiple functions is challenging when the storage system is distributed and has no central expert. The dispersed storage system not only supports secure and robust data storage and retrieval, but also lets a user frontward his data in the storage servers to another user without retrieving the data back.Using Cloud Storage, users can remotely store their data and enjoy the on-demand high quality applications and services from a shared pool of configurable computing resources, without the burden of local data storage and maintenance. However, the fact that users no longer have physical possession of the outsourced data makes the data integrity protection in Cloud Computing a formidable task, especially for users with constrained computing resources. Moreover, users should be able to just use the cloud storage as if it is local, without worrying about the need to verify its integrity. Thus, enabling public auditability for cloud storage is of critical importance so that users can resort to a third party auditor (TPA) to check the integrity of outsourced data and be worry-free. To securely introduce an effective TPA, the auditing process should bring in no new vulnerabilities towards user data privacy, and introduce no additional online burden to user. In this paper, we propose a secure cloud storage system supporting privacy-preserving public auditing. We further extend our result to enable the TPA to perform audits for multiple users simultaneously and efficiently. Extensive security and performance analysis show the proposed schemes are provably secure and highly efficient [5] several anonymization techniques, such as generalization and bucketization, have been designed for privacy preserving microdata publishing. Recent work has shown that generalization loses considerable amount of information, especially for high-dimensional data. Bucketization, on the other hand, does not prevent membership disclosure and does not apply for data that do not have a clear separation between quasi identifying attributes and sensitive attributes. In this paper, we present a novel technique called slicing, which partitions the data both horizontally and vertically. We show that slicing preserves better data utility than generalization and can be used for membership disclosure protection. Another important advantage of slicing is that it can handle high-dimensional data. We show how slicing can be used for attribute disclosure protection and develop an ef- ficient algorithm for computing the sliced data that obey the ℓ-diversity requirement. Our workload experiments confirm that slicing preserves better utility than generalization and is more effective than bucketization in workloads involving the sensitive attribute. Our experiments also demonstrate that slicing can be used to prevent membership disclosure.

III. PROPOSED ARCHITECTURE

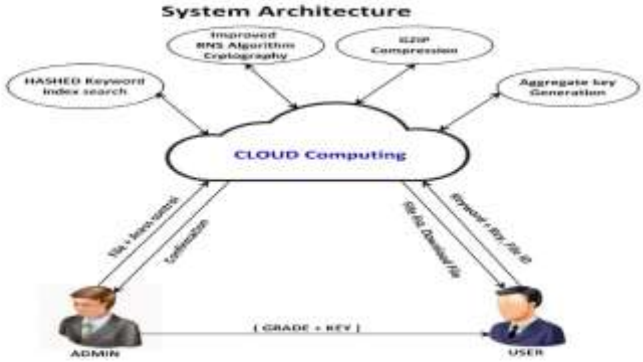

Fig 3: Overall System Architecture

The system architecture it consists of two modules admin and user. Admin upload the files to the cloud with access control and gets confirmation later in cloud the hashed keyword search, improved rns algorithm cryptography, GZIP compression and aggregate key generation takes place later user request the file to download with keyword and file id.

IV. METHODOLOGY AND ALGORITHM DESIGN

4.1 *Encryption and Decryption*

Considering data privacy, a traditional way to ensure it is to rely on the server to enforce the access control after authentication, which means any unexpected privilege escalation will expose all data. In a shared-tenancy cloud computing environment, things become even worse. Data from different clients can be hosted on separate virtual machines (VMs) but reside on a single physical machine. Data in a target VM could be stolen by instantiating another VM co-resident with the target one. Regarding availability of files, there are a series of cryptographic schemes which go as far as allowing a third-party auditor to check the availability of files on behalf of the data owner without leaking anything about the data, or without compromising the data owner's anonymity. Likewise, cloud users probably will not hold the strong belief that the cloud server is doing a good job in terms of confidentiality. A cryptographic solution, with proven security relied on number-theoretic assumptions is more desirable, whenever the user is not perfectly happy with trusting the security of the VM or the honesty of the technical staff. These users are motivated to encrypt their data with their own keys before uploading them to the server.

Data sharing is an important functionality in cloud storage. For example, bloggers can let their friends view a subset of their private pictures; an enterprise may grant her employees access to a portion of sensitive data. The challenging problem is how to effectively share encrypted data. Of course users can download the encrypted data from the storage, decrypt them, then send them to others for sharing, but it loses the value of cloud storage. Users should be able to delegate the access rights of the sharing data to others so that they can access these data from the server directly. However, finding an efficient and secure way to share partial data in cloud storage is not trivial.

1.2 RNS algorithm

1. Key Generation

Generate two large prime numbers, *p* and *q*

Let $n = pq$

Let $m = (p-1)(q-1)$

Choose a small number *e*, coprime to *m*

Find *d*, such that $de \% m = 1$

Publish e and n as the public keyKeep *d* and *n* as the secret key.

2. Encryption

$C = Pe \% n$

3. Decryption

$P = Cd \% n$

$x \% y$ means the remainder of *x* divided by *y*

4.3 Improved RNS Algorithm:

First, we have to select two primary keys.

Consider,

P1 = 11,

P2 = 13

Data N = 80

Key Generation:

M = P1 * P2 =143

A1 = M / P1 = 143 / 11 = 13

A2 = M / P2 = 143 / 13 = 11

T Value is, it can be anything

T1 = ((A1 * T) mod P1) == 1

T1 = 6

T2 = ((A2 * T) mod P2) == 1

T2 = 6

Encryption Process:

R1 = N % P1 = 80 % 11 = 3

R2 = N % P2 = 80 % 13 = 2

Decryption Process:

E = [(A1 * T1 * R1) + (A2 * T2 * R2)] mod M

E = [(13 * 6 * 3) + (11 * 6 * 2)] mod 143

E = [234 + 132] mod 143

E = [366] mod 143

E = 80

The Improved RNS Key Exchange Algorithm one of the most widespread and commonly used methods and intriguing in the distribution and exchange of key. It can be said that this algorithm is a public key encryption system; the only objective of this algorithm is key distribution.

This algorithm can be regarded as an example of a public key distribution scheme. In fact, we can say that the important feature of this algorithm, the exchange of a single piece of information, where the value getting is used as a session key for a private-key scheme.

V. RESULT ANALYSIS

TABLE 1

Tabulated data time taken by RNS and improved RNS system to download the files with the file size mentioned.

File Size	RNS in sec	Improved RNS in sec
20	91	87
40	93.5	87.5
60	96	90
80	98.5	94.5
100	101	96
120	103.5	100.5
140	106	101
160	108.5	106.5
180	111	109
200	113.5	112.5
220	116	114
240	118.5	115.5
260	121	117
280	123.5	118.5

The above table shows the time taken by RNS and improved RNS system to download the files with the file size mentioned.The Tabular data clearly says that based on the different file size, the time taken to download files in the improved RNS system is less when compared to RNSsystem

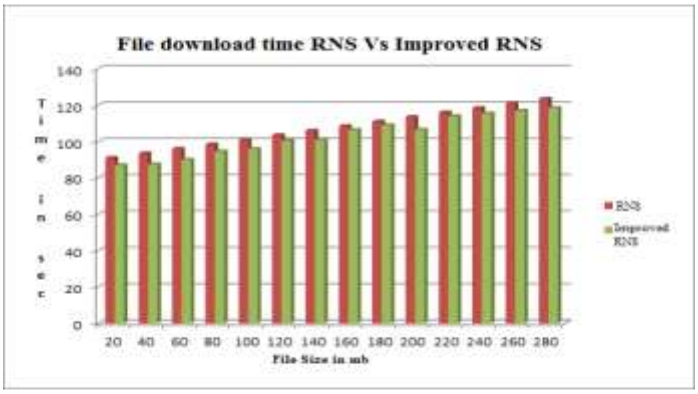

Fig 5.1: Comparison of File Download Time RNS Vs Improved RNS

The above graph shows the comparison of file downloading time between RNS system and the improved RNSsystem. X-axis shows the different file sizes in MB and Y-axis shows the corresponding time required for downloading those particular files in sec.The graph plotted above clearly shows that the time required to download the files in improved RNSsystem is less when compared to RNS system.

TABLE 2

Tabulated data time taken by RNS and improved RNS system to encrypt the files with the file size mentioned.

File Size	RNS in sec	Improved RNS in sec
20	160	91
40	150.5	93.5
60	156	96
80	167.5	98.5
100	148	101
120	176.5	103.5
140	181	106
160	165.5	108.5
180	158	111
200	185.5	113.5
220	159	116
240	173.5	118.5
260	169	121
280	191.5	123.5

The above table shows the time taken by RNS and improved RNS system to encrypt the files with the file size mentioned.The Tabular data clearly says that based on the different file size, the time taken to encrypt the files in the improved RNS system is less when compared to RNS system. When we consider the maximum file sizes we can save the more time to encrypt the file subsequently it improve the system performance. The following graph showed the comparison.

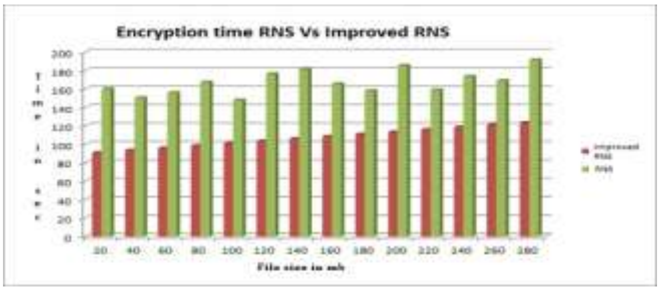

Fig 5.2: Comparison of File Encryption Time RNS Vs Improved RNS

The above graph shows the comparison of file encryption time between RNS and the improved RNS system. X-axis shows the different file sizes in MB and Y-axis shows the corresponding time required for encrypting those particular files in sec.The graph plotted above clearly shows that the time required to encrypt the files in improved RNSis less when compared to the RNS system.

TABLE 3

Tabulated data time taken by RNS and improved RNS system to decrypt the files with the file size mentioned.

File Size	RNS in sec	Improved RNS in sec
20	8	6
40	9.5	6.5
60	11	7
80	12.5	7.5
100	14	8
120	15.5	8.5
140	17	9
160	18.5	9.5
180	20	10
200	21.5	10.5
220	23	11
240	24.5	11.5
260	26	12
280	27.5	12.5
300	29	13

The above table shows the time taken by RNS and improved RNS system to decrypt the files with the file size mentioned.The Tabular data clearly says that based on the different file size, the time taken to decrypt files in the improved RNS is less when compared to RNS system. When we consider the maximum file sizes we can save the more time to encrypt the file subsequently it improve the system performance. The following graph showed the comparison.

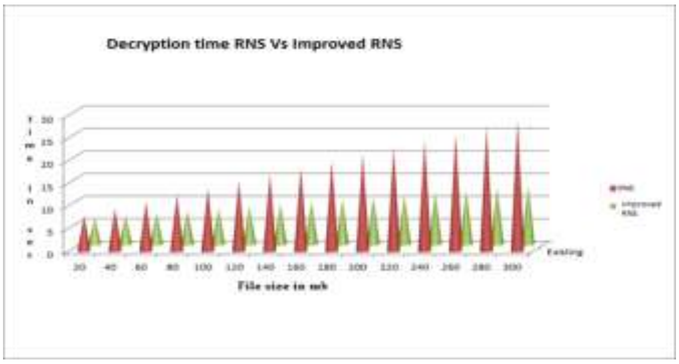

Fig 5.3: Comparison of File Decryption Time RNS Vs Improved RNS

The above graph shows the comparison of file decryption time between RNS and the improved RNS system. X-axis shows the different file sizes in MB and Y-axis shows the corresponding time required for decrypting those particular files in sec.The graph plotted above clearly shows that the time required to decrypt the files in improved RNS is less when compared to the RNS system.

VI. CONCLUSION AND FUTURE WORK

This system developed in MVC architecture implemented and tested in hybrid cloud approach. The experimental results show the system meet the all designed constraints which are discussed in system analysis.The system automatically extract keywords from uploading file and the keywords are converted into hash key with respective access control while users are searching for a file by providing keywords based on users grade hash key will be generated and searched in hashkey index and corresponding files are retrieved for security and efficiency.

In this proposed system we are using cloud for storage, as a future enhancement. We can use Hadoop for storage system and we can increase searching efficient with more number of keywords.

REFERENCES

[1] “Data Partitioning Technique to Improve Cloud Data Storage Security,” Swapnil V.Khedkar , A.D.Gawande; International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 3347-3350.

[2] “cloud computing based rural e-governance model”, ashish bhushan khare, vishal raghav and prateek sharma, Journal of Information and Operations Management, ISSN: 0976–7754 & E-ISSN: 0976–7762 , Volume 3, Issue 1, 2012, pp-89-91.

[3] “Privacy- Preserving Public Auditing for Secure Cloud Storage,” C. Wang, S.S.M. Chow, Q. Wang, K. Ren, and W. Lou, IEEE Trans.Computers, preprint, 2012.

[4] "A Secure Erasure Code-Based Cloud Storage System with Secure Data Forwarding," Hsiao-Ying Lin; Tzeng, W.-G.; Parallel and Distributed Systems, IEEE Transactions on , vol.23, no.6, pp.995-1003, June 2012.

[5] "HASBE: A Hierarchical Attribute-Based Solution for Flexible and Scalable Access Control in Cloud Computing," Zhiguo Wan; Jun'e Liu; Deng, R.H.; Information Forensics and Security, IEEE Transactions on , vol.7, no.2, pp.743-754, April 2012.

[6] "Enabling Public Auditability and Data Dynamics for Storage Security in Cloud Computing," Qian Wang; Cong Wang; Kui Ren; Wenjing Lou; Jin Li; , Parallel and Distributed Systems, IEEE Transactions on , vol.22, no.5, pp.847-859, May 2011.

[7] “Mapping Cloud Computing onto Useful eGovernance”, Ajay Prasad, Sandeep Chaurasia, Arjun Singh, Deepak Gour, (IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No. 5, 2010.

[8] “Cloud Computing: Future Framework for e-Governance”, K.Mukherjee, G.Sahoo, International Journal of Computer Applications (0975 – 8887)Volume 7– No.7, October 2010.

[9] "A New Architecture of a Two-Stage Lossless Data Compression and Decompression Algorithm," Ming-Bo Lin; Yung-Yi Chang, Very Large Scale Integration (VLSI) Systems, IEEE Transactions on , vol.17, no.9, pp.1297,1303, Sept. 2009.

[10] "Huffman and Lempel-Ziv based data compressi20on algorithms for wireless sensor networks," Renugadevi, S.;

Nithya Darisini, P.S.,Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013

[11] "Privacy-Preserving Public Auditing for Secure Cloud Storage," Cong Wang; Chow, S.S.M.; Qian Wang; Kui Ren; Wenjing Lou,Computers, IEEE Transactions on , vol.62, no.2, pp.362,375, Feb. 2013.

[12] “Cloud Computing: Solving Availability Problem in Future Framework for e-Governance”, Dileep Kumar Gupta Abhishek Mishra Dr. G. Sahoo, International Journal of Computer Applications & Information Technology Vol. 2, Issue II Feb-March 2013.

[13] “Hybrid Cloud Computing in E-Governance: Related Security Risks and Solutions”, Hardayal Singh Shekhawat and D.P. Sharma, Research Journal of Information Technology 4(1): 1-6, March 10, 2012.

[14] ‘Slicing: A New Approach for Privacy Preserving Data Publishing," Tiancheng Li; Ninghui Li; Jian Zhang; Molloy, I.;Knowledge and Data Engineering, IEEE Transactions on , vol.24, no.3, pp.561-574, March 2012.

[15] "Toward Secure and Dependable Storage Services in Cloud Computing," Wang Cong, Wang Qian, Ren Kui, Cao Ning and Lou Wenjing , Services Computing, IEEE Transactions on , vol.5, no.2, pp.220-232, April/June 2012

AUTHOR PROFILE

[1]**Dr. Bondu Venkateswarlu** is with faculty of Computer Science and Engineering, Dayananda Sagar University, Bangalore, India. He received the Ph.D Degree in Computer Science and Systems Engineering from Andhra University College of Engineering, AndhraUniversity. He is field of specialization in Data Mining and his research areas of interests are Data Analytics, Software Engineering & Data Modeling. He is a life member of ISTE and academic nominee member for CSI.

[2]**Soumya G V**, M.Tech in Computer science and information technology specialization in cloud computing Dayananda sagar university Bangalore. Her research area is Big data and cloud computing and cloud security