iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 5 **Issue:** III **Month of publication:** March 2017

**DOI:** http://doi.org/10.22214/ijraset.2017.3047

www.ijraset.com

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

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

# Implementation on Secure Distributed Deduplication Systems with Improved Reliability

Ms. Priyanka Jerpoth[1], Ms. Leena Shirpurkar[2], Ms. Aachal Kamble[3], Prof. Vaishali Sarangpure[4]
[1,2,3]*Final Year B.E Dept. of CSE, SRMCEW, RTMNU, Nagpur, India.*
[4]*Professor, B.E. Dept. Of CSE, SRMCEW, RTMNU, Nagpur, India.*

***Abstract*: *Data deduplication is a method for removing multiple copies of same data, and has been widely used in cloud storage to reduce storage space and uploading speed. However, there is only one copy for each file stored in cloud even if such a file is owned by a multiple number of users. As a result, deduplication system improves storage utilization while reducing reliability. Mostly, the challenge of privacy for sensitive data also arises when they are outsourced by users to cloud. Aiming to address the above security challenges, this paper makes the first attempt to formalize the notion of reliable deduplication system. We propose new distributed deduplication systems with higher reliability in which the data chunks are distributed across multiple cloud servers. The data confidentiality and tag consistency of data are also achieved by introducing a deterministic secret sharing scheme in distributed storage systems, instead of using convergent encryption as in previous deduplication systems. Security analysis depicts that our deduplication systems are secure in terms of the definitions specified in the proposed security model. As a proof of concept, we implement the proposed systems and demonstrate that the incurred overhead is very limited in realistic environments.***

***Keywords:Deduplication, Reliability, Distributed System.***

## I. INTRODUCTION

With the huge growth of digital data, deduplication techniques are widely employed to backup data and minimize network and storage overhead by detecting and eliminating redundancy among data. Instead of keeping multiple data copies with the same content, deduplication eliminates redundant data by keeping only one physical copy and referring other redundant data to that copy. Deduplication has received much attention from academics and industry because it can greatly improves storage utilization and save storage space, especially for the applications with high deduplication ratio such as storage systems. A number of deduplication systems have been proposed based on various deduplication strategies such as client-side or server-side deduplications, file-level or block-level deduplications. A brief review is given in Section 6. Especially, with the advent of cloud storage, data deduplication techniques become more attractive and critical for the management of ever-increasing volumes of data in cloud storage services which motivates enterprises and organizations to outsource data storage.To third-party cloud providers, as evidenced by many real-life case studies . According to the analysis report of IDC, the volume of data in the world is expected to reach 40 trillion gigabytes in 2020 . Today's commercial cloud storage services, such as Drop box, Google Drive. have been applying deduplication to save the network bandwidth and the storage cost with client-side deduplication. There are two methods of deduplication in terms of the size: (i) file-level deduplication, which discovers redundancies between different files and removes these redundancies to reduce capacity demands, and (ii) block level deduplication, which discovers and removes redundancies between data blocks. The file can be divided into smaller fixed-size or variable-size blocks. Using fixedsize blocks simplifies the computations of block boundaries, while using variable-size blocks (e.g., based on Rabin fingerprinting [3]) provides better deduplication efficiency. Though deduplication technique can save a large amount of storage space for the cloud storage service providers, it reduces the reliability of the system. Data reliability is actually a very critical issue in a deduplication storage system because there is only one part of each file stored in the server shared by all the owners. If such a shared file/chunk was lost, a disproportionately large amount of data becomes inaccessible because of the unavailability of all the files that share this file/chunk. If the value of a file were measured in terms of the amount of file data that would be lost in case of losing a single file, then the amount of user data lost when a file in the storage system is corrupted

## II. RELATED WORK

Cloud computing is an Internet based development of computer technology. It is a model that allow convenient, on- demand network access to a shared pool of various customizable computing resources. The term Cloud refers to a Network or an Internet. In

other way, we can say that it is something, which is present at remote location. Cloud can provide various services over network, such as on public networks and on private networks, viz., WAN, LAN or VPN. Various applications such as an e-mail, web conferencing, customer relationship management (CRM), all run on Cloud.Cloud computing combine virtualization, on-demand deployment, Internet delivery of services, and open source software. Cloud computing uses various approaches, concepts, and best practices that have been established. Everything is new because cloud computing changes on how we invent, develop, deploy, scale, update, maintain, and pay for applications and the infrastructure on which they run.

Cloud computing is architecture for providing computing service via the internet on demand and pay per use access to a pool of shared resources viz. various types of networks, storage, servers, services and applications, without physically acquiring. So it saves managing cost and time for organizations. In present days cloud computing is one of the greatest platform that provides storage of data at very low cost and is available at any time over the internet. But it has various critical issue in security, load management and fault tolerance.Data sharing refers to storing data at a place where it can use by multiple users, at the same time ensuring the security of data. In this project data is shared using a secured mechanism that makes use of encryption for ensuring security of data as well as it also contains mechanisms for authentication of users. The different security mechanisms used in this project are encryption and compression.

## III. PROPOSED ALGORITHM

In recent years, cloud computing has gained considerable attention in the industry, academia, and the public sector. It can be seen as a realization of the utility computing paradigm that establishes the notion of computing services that are delivered on demand. Cloud computing provides infrastructure, platform and software services over a network connection. For example: It allows rapidly developing and provisioning applications on remote servers due to pre-built software platform that are offered, operated and maintained by cloud provider. Cloud computing combines several existing technologies to form a new computing paradigm, for example, it relies on a broad network access and most often employs virtualization technology. It is the combination of those technologies that facilitates the emergence of a central new characteristics of cloud computing as it allows to dynamically and rapidly scale resource based on the actual user demand. Today the Computer industry fails to seize the vast commercial potential of cloud computing. Instead, cloud computing catalyses' the pressure on networking. Cloud offers business opportunities to perform things faster and better:

Scalability: Data and application resources can be provided quickly when and where it is needed.

Availability: With the right cloud provider, we can make sure that the resources remain continuously available and are always kept secured.

Less Maintenance: Hardware, applications and bandwidth are managed by the provider and hence less maintenance is required.

Expert service: For our convenience cloud computing services are continuously monitored and maintained by expert data center technicians.

## IV. CONNECTING WEBSITE TO CLOUD DATABASE

This section provides a sample script code that creates a simple webpage. You can use this webpage to test that whether your MySQL database is working or not. You can also use it as a simple calculator. You should copy the script and paste it into any of the text editor. Then you can modify the script by using your own hostname, user name, password, and database instance name information and then make sure to save the changes. Finally, you must copy this script to your cloud server and then execute the script to display the simple webpage and then test your connection of your website to your database instance. Make sure that your web server must be in the same region as your database instance.

## V. FILE ENCRYPTION AND SPLITTING

If any of the file contains very sensitive information, then you can encrypt the file while compressing it. Option -e encrypts the file with the given password, and the receiver should know this password for decrypting it. If the size of file exceeds the specified limit after compressing also, then we must split the files into smaller sizes.

## VI. REMOVING DUPLICATIONS AND TESTING

Removing duplication from the files means that the repeated data should be deleted so that the space is not wasted and this space will be made available for another purpose. So the less space will be required and thus without wasting the space another task can be

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

perform with that space and after that twisting is done.

## VII. AES ALGORITHM

AES is based on a design principle known as a Substitution and permutation network. It is faster in both software and hardware. Unlike its predecessor viz., DES, AES does not make use of a Feistel network. AES has a fixed block size of 128 bits and varying key size of 128, 192, or 256 bits, whereas Rijndael's can be specified with block and key sizes in any multiple of 32 bits, with a minimum of 128 bits.The block size has a maximum number of 256 bits, but the key size has no theoretical maximum. AES operates on a 4×4 column-major order matrix of bytes, termed as the state (versions of Rijndael's with a larger block size have additional columns in the state). Most of the AES calculations are done in a special finite field.

Working of AES:Advanced Encryption Standard or AES was invented by "J. Daemen and V. Rijmen", and is successfully accepted by the US federal government in the year 2001 for the top secret approved encryption algorithms. It is also referred to as Rijndael's, as it is based on the Rijndael algorithm. Reportedly, this standard has never been cracked.AES has the following three approved key size: 128 bits, 192 bits, and 256 bits. Here we are trying to explain the working process of the algorithm in simple terms: an algorithm starts by selecting a random number, in which the key and data encrypted with it are mixed using four different rounds of mathematical processes. The key that is used to encrypt the number must also be used to decrypt it.The four rounds are called as the Sub-Bytes step, the Shift-Rows step, the Mix Columns step, and the last is the AddRoundKey step. During the execution of Sub-Bytes round, a lookup table is used to determine what each byte is replaced with. The Shift- Rows step has some number of rows where each row of the state is shifted cyclically by a particular offset value, while leaving the first row unchanged. Each byte of the second row is shifted to the left, by an offset of value one, each byte in the third row by an offset value of two, and the fourth row by an offset value of three and so on. This shifting process is applied for all three key lengths, whereas there is a slight difference for the block size 256-bits where the first row is kept unchanged, the second row is shifted by an offset value of one, the third by an offset value of three, and the fourth by an offset value of four. The Mix Columns step of the algorithm is the mixing operation using linear transformation in order to combine the four bytes in each column. The four bytes are taken as input and output the output is generated.

In the fourth round, the AddRoundKey derives round keys from Rijndael's key schedule, and adds the round key to each byte of the state. Each round key is added by combining each byte of the state with the corresponding byte from the round key. Lastly, these steps are repeated again for the fifth round, but the only difference is that it does not include the Mix Columns step.This algorithm takes the basic data known as plain text as an input and changes it into a code known as cipher text. The larger the key size, the greater the number of potential patterns that can be is reason why AES has been Teflon-coated.

## VIII. CONCLUSION AND FUTURE WORK

As we have mentioned above, the major requirement is balancing of the work load of database server so that it is easy to interfere with cloud. In this project, all these issues are recovered by the splitting technique. In our system, firstly the user interacts with the web browser which delegates the servers scripting or server's activities. For storing the data on cloud, it is necessary to have one account on particular cloud therefore we are providing a facility to create one account by which appropriate user will interact with the data. It includes sharing of data using uploading and downloading functionality and also secures it using AES algorithm. This project also covers various features like security through username and password, user can select the file and default mode of upload. This project would be designed to perform better than other file sharing tools available.

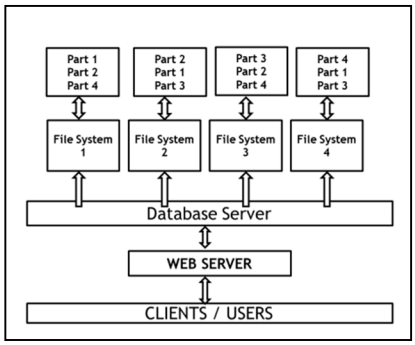


Fig.1: Proposed System Architecture

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

## IX. RESULT OF THE SYSTEM

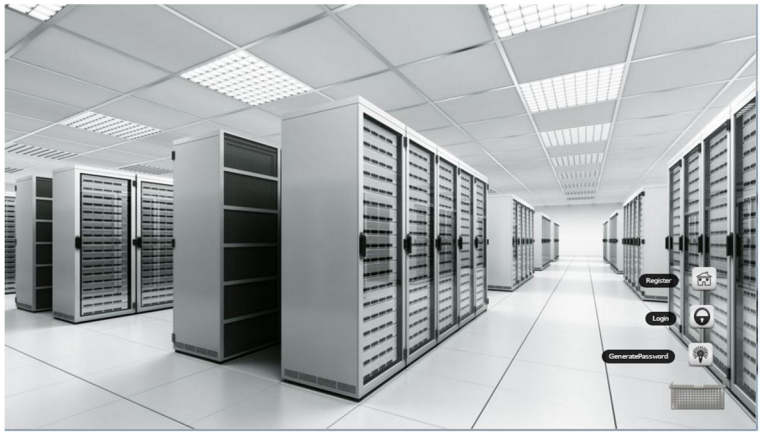



FIG 2: HOME PAGE

When the user starts the system the above page is displayed to the user. It contains the three option : a. Register, b. Login, c. Generate password.


FIG 3: USER AUTHENTICATION

Basically whenever a user wants to use the system he/she is required to register onto the system if not registered. After registration the email is verified by sending the temporary password on mail itself. Ones the user has id and password he can login into the system and use system services.


FIG 4:NEW USER REGISTRATION

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

For using the cloud services, it must to register first into cloud with valid email-id for getting the temporary generated password. In that registration form user put his/ her valid information. After registration, temporary password is send to that email-id which is used for login and used cloud services. Here we also provide to select the profile picture by which user authentication is indicated.


Fig 5: Login window

After the new user has registered himself into the system then by entering the email id and password he/she must have to login into the system.

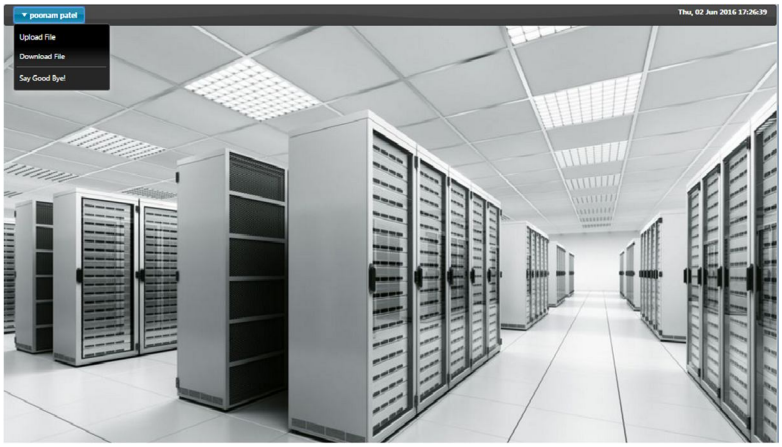



Fig 6: Successful login

After the user has successfully login into the system the the above window is displayed to the user. The window contains three menus : a. upload file , b. download file, c. Say Good bye (logout).

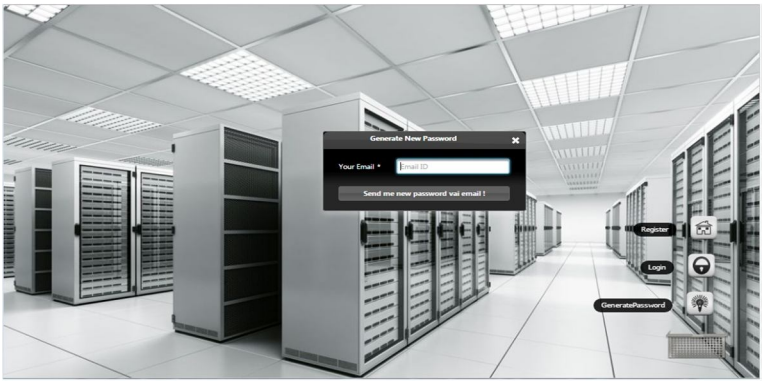



Fig 7: Forgot password

Using this form user can generate a new password for provided email id. The system generates a new password for the given mail id

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

and then sends it to user email id.

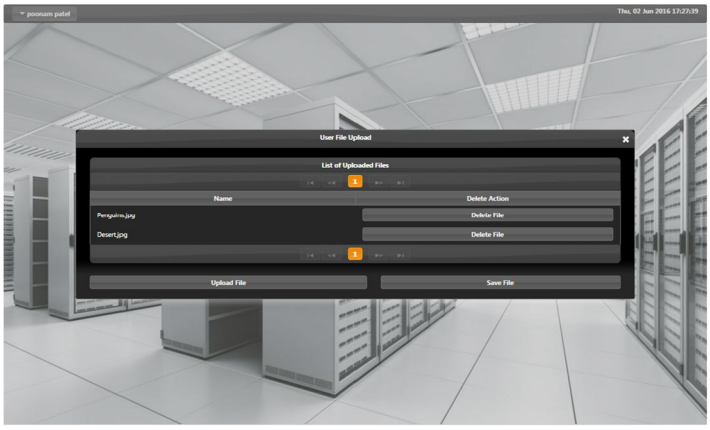



Fig 8: File Upload

Once user logged in user can access the services of cloud like uploading. In above figure we can see that the dialog box of uploading the file. After clicking on upload file, it's must to save that file. After uploading, the file will be splitted into three cloud database by which system overcome load balancing issue. In that system, user can upload any type of file like jpeg, bmp, mp3, mp4, txt, .avi, .mkv, etc.

Whenever a file is uploaded to server the web server it first encrypts the file the compress it and at last the file gets splitted into 4 equal parts and is stored in three different databases and revert is done while downloading.

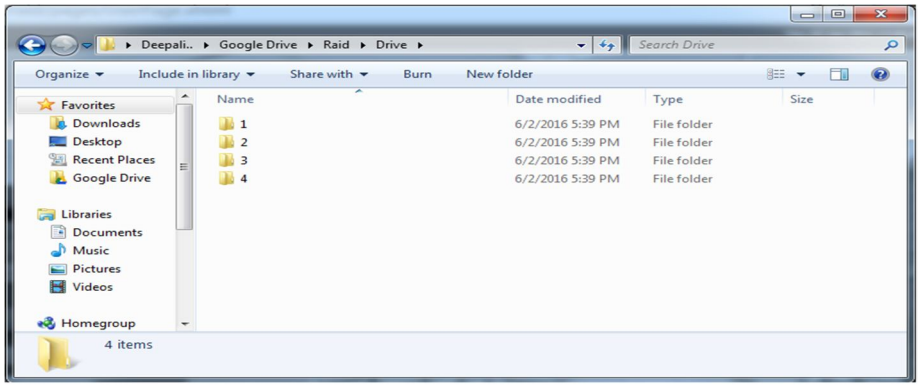



Fig 9: Four Different Cloud Databases

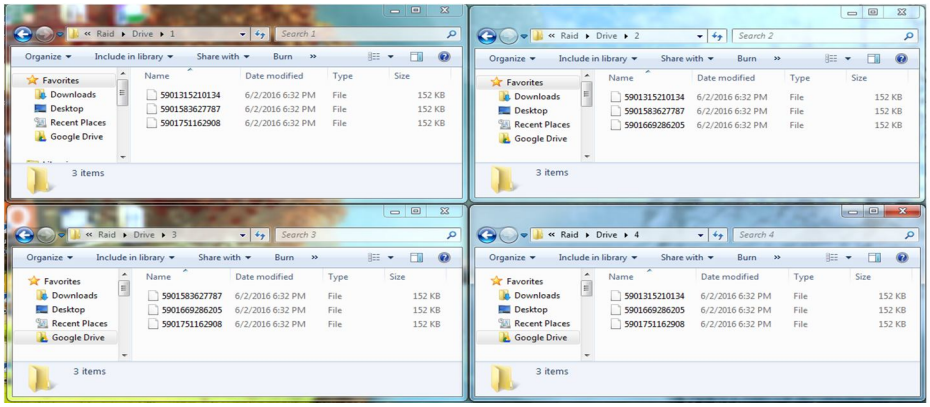


Fig 10: File Merge / Split

Volume 5 Issue III, March 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

*A. Download File*

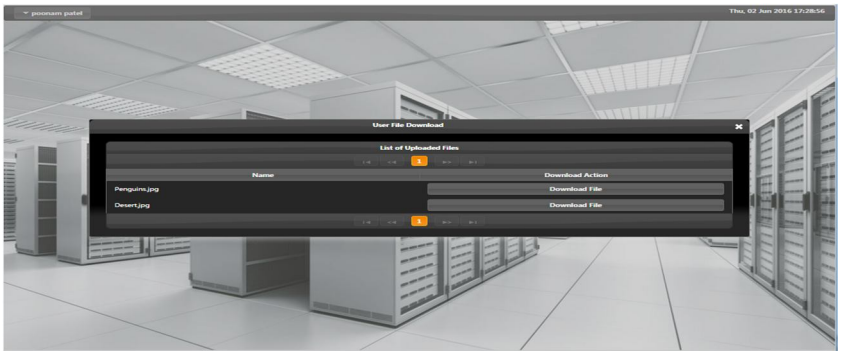



Fig 11: Download File

Above window will be displayed when the user selects the option to download the file. Here user can download his file by

*B. Database Design*

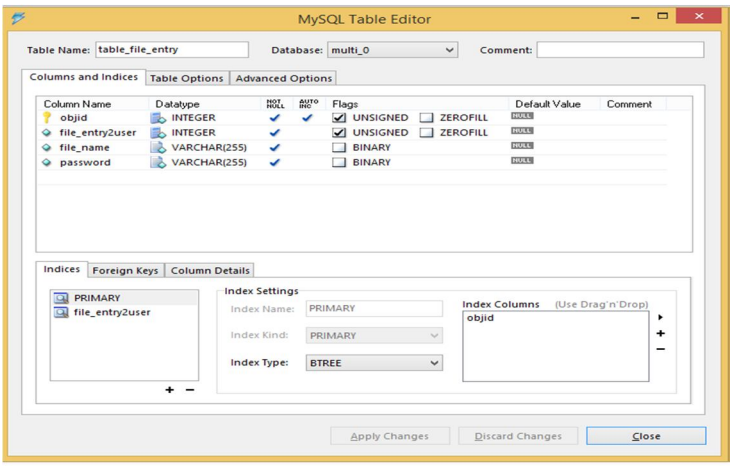



Fig: 6.11 File Entry

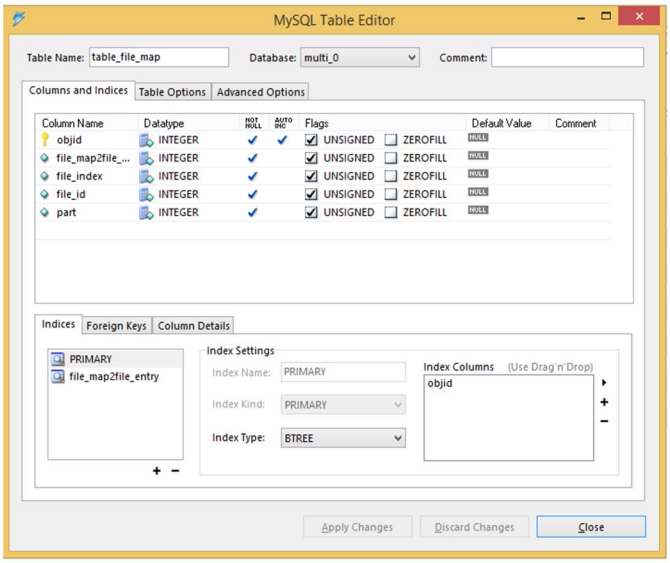


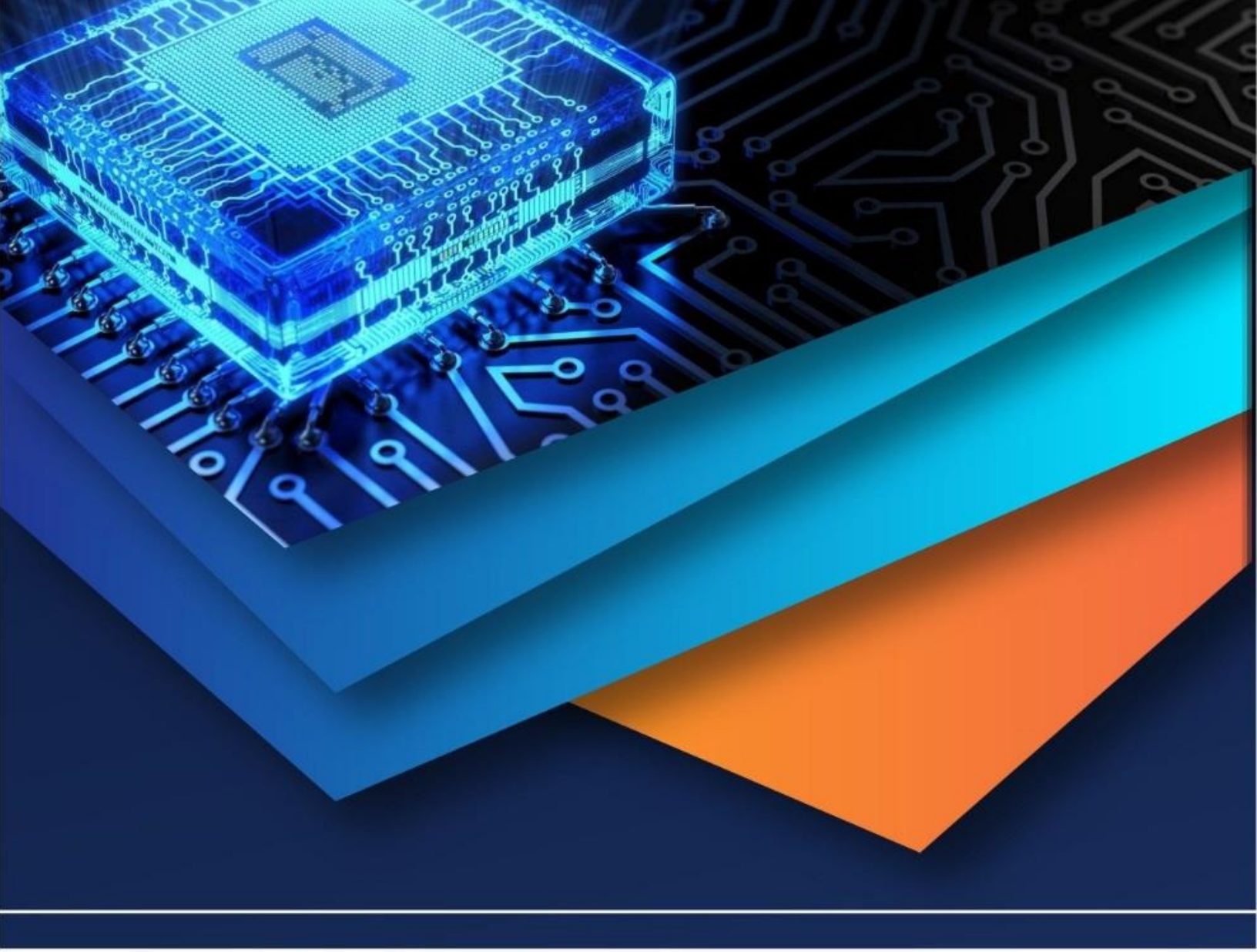

Fig 13: File Mapping

## X. CONCLUSION

This Project implements the distributed de-duplication systems to improve the Reliability of data while achieving the Confidentiality of the users' outsourced data without an encryption mechanism. It support file-level and fine-grained block-level data de-duplication. The security of tag consistency and integrity is achieved. We implemented our de-duplication systems using the Ramp secret sharing scheme and demonstrated that it incurs small encoding/decoding overhead compared to the network transmission overhead in regular upload/download operations.

## REFERENCES

[1] Mohammed A. Alzain , Eric Pardede , Ben Soh , James A. Thom "Cloud Computing Security: From Single To Multi-Clouds", 45th Hawaii International Conference On System Sciences 2012.
[2] OlfaNasraoui, Member, IEEE, MahaSoliman, Member, IEEE, EsinSaka, Member, IEEE, Antonio Badia, Member, IEEE, And Richard Germain "Ensuring Data Integrity And Security In Cloud Storage", IEEE TRANSACTIONS ON CLOUD AND DATA ENGINEERING, VOL. 20, No. 2, February 2013.
[3] Qin Liu ,ChiuC.Tan ,Jiewu, And Guojun Wang "Reliable Re-Encryption In Unreliable Clouds", IEEE Communications Society Subject Matter Experts For Publication In The IEEE Globecom 2011 Proceedings.
[4] Wei-Tek Tsai, Xin Sun, JanakaBalasooriya "Service-Oriented Cloud Computing Architecture", 2010 Seventh International Conference On Information Technology
[5] Cheng-Kang Chu, Sherman S. M. Chow, Wen-GueyTzeng, Jianying Zhou, And Robert H. Deng, Senior Member, IEEE "Key-Aggregate Cryptosystem For Scalable Data Sharing In Cloud Storage", IEEE Transactions On Parallel And Distributed Systems. Volume: 25, Issue: 2. Year: 2014
[6] Mell-Peter, Grance-Timothy "The NIST Definition Of Cloud Computing", September 2011.
[7] C. Cachin, I. Keidar And A. Shraer "Trusting The Cloud", ACM SIGACT News, 40, 2009, Pp. 81-86. Clavister, "Security in The Cloud", Clavister White Paper, 2008.
[8] H.Mei, J. Dawei, L. GuoliangAnd Z. Yuan "Supporting Database Applications As A Service", ICDE'09:Proc. 25thintl.Conf. On Data Engineering, 2009, Pp. 832-843.C. Wang, Q. Wang, K. Ren and W. Lou, "Ensuring Data Storage Security In Cloud Computing", ARTCOM'10: Proc. Intl. Conf. On Advances In Recent Technologies



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY