iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 4 **Issue:** X **Month of publication:** October 2016

DOI:

www.ijraset.com

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

Volume 4 Issue X, October 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

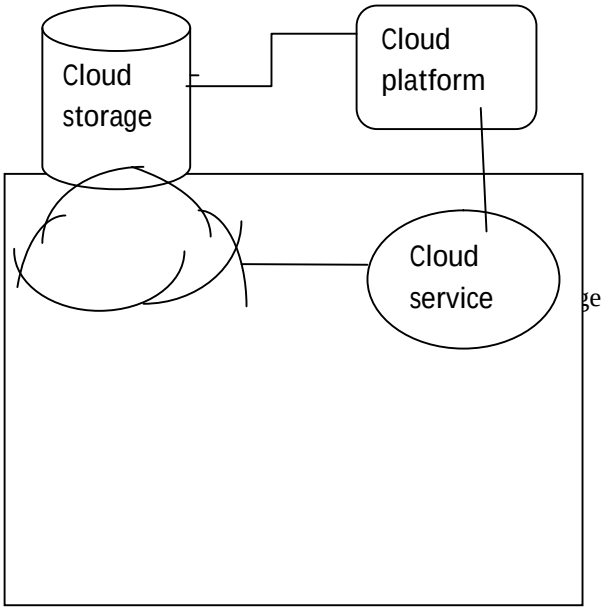
Reconciling End-To-End Confidentiality and Data Reduction in Cloud Storage

Sreedarsa K S[1], T Sivakumar[2], Neethu Rosiline[3]
Computer Science and Engineering, Anna University, Coimbatore
[1] *PG Student, Maharaja Institute of Technology, Coimbatore,*
[2] *M.E.(CSE), Professor, Maharaja Institute of Technology, Coimbatore,*
[3] *PG Student, Maharaja Institute of Technology, Coimbatore.*

Abstract- The repeating of data storage is reduced by the implementation of the de-duplication techniques. The wastage of memory storage is affected the storage space. The concept of cloud computing is helpful for data storage. But the duplication is possible; this will affected the storage capacity. In the Cloud computing platform using some comparison checking with some privileges, then increase the storage space utilization and increase the security of data file.
Keywords: de-duplication data, convergent encryption, duplicate check.

I. INTRODUCTION

Cloud computing provides seemingly unlimited "virtualized" resources to users as services across the whole Internet, while hiding platform and implementation details. Today's cloud service providers offer both highly available storage and massively parallel computing resources at relatively low costs. As cloud computing becomes prevalent, an increasing amount of data is being stored in the cloud and shared by users with specified privileges, which define the access rights of the stored data. One critical challenge of cloud storage services is the management of the ever-increasing volume of data. To make data management scalable in cloud computing [21], de-duplication has been a well-known technique and has attracted more and more attention recently. Data de-duplication is a specialized data compression [18] technique for eliminating duplicate copies of repeating data in storage. The technique is used to improve storage utilization [7] and can also be applied to network data transfers to reduce the number of bytes that must be sent. Instead of keeping multiple data copies with the same content, de-duplication[1,3,5,18] eliminates redundant data by keeping only. One physical copy and referring other redundant data to that copy. De-duplication [2,5,13,14,16] can take place at either the file level or the block level. For file-level de-duplication [17], it eliminates duplicate copies of the same file. De-duplication [13] can also take place at the block level, which eliminates duplicate blocks of data that occur in non-identical files.



A. Related Work

Title: Fast and Secure Laptop Backups with Encrypted De-duplication
Author: Paul Anderson, Le Zhang
Year : 2011
Description: Many people now store large quantities of personal and corporate data on laptops or home computers. These often have poor or intermittent connectivity, and are vulnerable to theft or hardware failure. Conventional backup solutions are not well suited to this environment, and backup regimes are frequently inadequate. This paper describes an algorithm which takes advantage of the data which is common between users to increase the speed of backups, and reduce the storage requirements. This algorithm supports client-end per-user encryption which is necessary for confidential personal data. It also supports a unique feature which allows immediate detection of common sub trees, avoiding the need to query the backup system for every file. We describe a prototype implementation of this algorithm for Apple OS X, and present analysis of the potential effectiveness, using real data obtained from a set of typical users. Finally, we discuss the use of this prototype in conjunction with remote cloud storage, and present an analysis of the typical cost savings.

Title: Clouded up: secure de-duplication with encrypted data for cloud storage
Author: Pasquale puziosecludit and eurecom, refikmolvaeurecom, melek oneneurecom
Year : 2012
Description: With the continuous and exponential increase of the number of users and the size of their data, data de-duplication becomes more and more a necessity for cloud storage providers. By
storing a unique copy of duplicate data, cloud providers greatly reduce their storage and data transfer costs. The advantages of de-duplication unfortunately come with a high cost in terms of new security and privacy challenges. We propose Cloud De-duplication, a secure and efficient storage service which assures block-level de-duplication and data confidentiality at the same time. Although based on convergent encryption, ClouDedup remains secure thanks to the definition of a component that implements an additional encryption operation and an access control mechanism. Furthermore, as the requirement for de-duplication at block-level raises an issue with respect to key management, we suggest including a new component in order to implement the management for each block together with the actual de-duplication operation.

Title: Secure de-duplication and data security with efficient and reliable CEKM
Author:N.O.agrawal, prof Mr. s.s.kulkarni
Year : 2014
Description: Secure de-duplication is a technique for eliminating duplicate copies of storage data, and provides security to them. To reduce storage space and upload bandwidth in cloud storage de-duplication has been a well-known technique. For that purpose convergent encryption has been extensively adopt for secure de-duplication, critical issue of making convergent encryption practical is to efficiently and reliably manage a huge number of convergent keys. The basic idea in this paper is that we can eliminate duplicate copies of storage data and limit the damage of stolen data if we decrease the value of that stolen information to the attacker. This paper makes the first attempt to formally address the problem of achieving efficient and reliable key management in secure de-duplication. We first introduce a baseline approach in which each user holds an independent master key for encrypting the convergent keys and outsourcing them. However, such a baseline key management scheme generates an enormous number of keys with the increasing number of users and requires users to dedicatedly protect the master keys. To this end, we propose Dekey, User Behavior Profiling and Decoys technology. Dekey new construction in which users do not need to manage any keys on their own but instead securely distribute the convergent key shares across multiple servers for insider attacker. As a proof of concept, we implement Dekey using the Ramp secret sharing scheme and demonstrate that Dekey incurs limited overhead in realistic environments. User profiling and decoys, then, serve two purposes: First one is validating whether data access is authorized when abnormal information access is detected, and second one is that confusing the attacker with bogus information. We posit that the combinations of these security features will provide unprecedented levels of security for the de-duplication in insider and outsider attacker.

Title: Identity-based identification and signature schemes using correcting codes
Author: Pierre-Louis Cayrel1, Philippe Gaborit1 and Marc Girault2
Year : 2011
Description: In this paper, we propose a new identity-based identification (and signature) scheme based on error-correcting codes. This scheme is up to date the first identity-based scheme not based on number theory. The scheme combines two well-known code-

based schemes: the signature scheme of Courtois, Finiasz and Sendrier and the zero-knowledge authentication scheme of Stern (which may also be used for signature).The scheme inherits from the characteristics of the previous schemes: it has a large public key of order 1Mo and necessitates a certain number of exchange rounds. The scheme can also work in signature but leads to a very large signature of size 1Mo.

Title: An Efficient Identification Scheme in Standard Model Based on the Diophantine Equation Hard Problem

Author: B.C. Tea, M.R.K. Ariffin and J.J. Chin

Year : 2013

Description: Recently the Diophantine Equation Hard Problem (DEHP) was proposed. It is utilized to design a standard identification scheme model. Since the computation involves only simple addition and multiplication steps, the efficiency and the time cost are greatly improved as compared to the existing identification schemes. In this paper, we propose a zero knowledge identification scheme based upon the DEHP. With the assumption such that DEHP is intractable, we provide the security analysis on the impersonation against non-adaptive passive attack (imp-pa) and show that our new proposed scheme is more desirable due to high efficiency in terms of time computation.

B. System Details

1) *Existing System:* In the existing system the de-duplication techniques not supporting authorization, duplicate checkup, which to be vital in several applications. The authorized de-duplication [4] system will be release a collection of privileges at the starting time. To save space, commercial cloud storage services such as Google drive [4] and DropBox [6] perform file level de-duplication across all their users. De-duplication[2,3,5,12,13], which is practical only with random-access devices, removes this redundancy by storing duplicate data[13] only once and has become an essential feature of disk-to-disk backup solutions. With the continuous and exponential increase of the number of users and the size of their data, data de-duplication becomes more and more a necessity for cloud storage providers [8]. To reduce storage space and upload bandwidth in cloud storage de-duplication has been a well-known technique [10].

   a) The existing de-duplication system, each user is issued a set of privileges during system initialization.

   b) Each file uploaded to the cloud is also bounded by a set of privileges to specify which kind of users is allowed to perform the duplicate check and access the files.

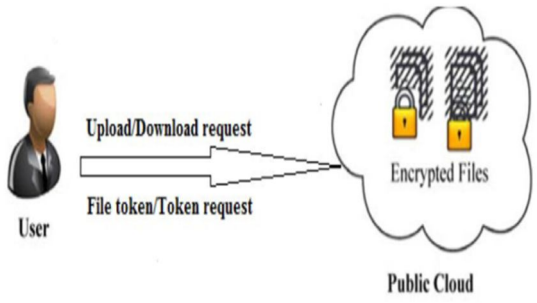
   c) Before submitting his duplicate check request for a file, the user needs to take this file and his own privileges as inputs. The user is able to find a duplicate for this file if and only if there is a copy of this file and a matched privilege stored in cloud.

C. Existing Technique

Existing System Technique Explanation

1) *Symmetric encryption technique:* Symmetric encryption uses a common secret key to encrypt and decrypt information. A symmetric encryption scheme consists of three primitive functions. The key generation algorithm that generates using security parameter. The symmetric encryption algorithm that takes the secret key and message and then outputs the cipher text and the symmetric decryption algorithm that takes the secret key and cipher text and then outputs the original message.

REF PAPER:"Server-aided encryption for DE duplicated storage"



b) Proposed System

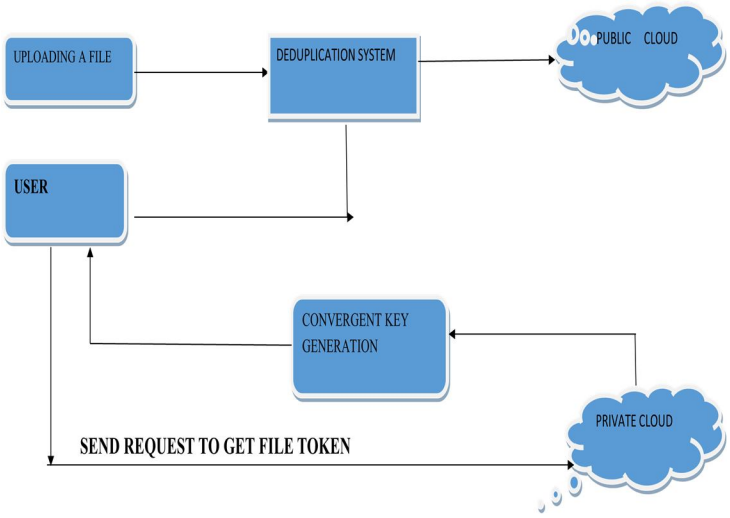
In the proposing system, eliminating duplicate copies [13, 22, 23] of repeating data and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the privacy of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing. To better protect data security [21, 22]; this paper makes the first attempt to formally address the problem of authorized data de-duplication [15, 16, and 17].

Proposed Technique: - Convergent encryption [19] technique

Technique Definition:-A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key**.**

Advantages:-The user doesn't needs to know private key. Better protect security [6, 10, 11].This de-duplication systems can support differential authorization [15, 16] duplicate check.

II. SYSTEM ARCHITECTURE



A. Design module

Four verity modules are present in this architecture. Authentication module [15, 16], Convergent Key [19] Generation module, File Uploading module, and Authorized [4] Duplicate Check Scheme module. The user first upon login for upload or download a data file along with the details of modules mentioned below,

1) *Authentication module:* The process of identifying an individual usually based on a username and password. In security systems, Authentication merely ensures that the individual is who he or she claims to be, but says nothing about the access rights of the individual. In authentication module is used to security purpose. Here this module only for user, after registration user enters the username and password. This input is check into the database, whether input is correct or not. If input is correct then allow to next process otherwise consider as a non authenticated user. In this Module If he is a new user he needs to enter the required data to register the form and the data will be stored in server for future authentication purpose.
2) *Convergent Key Generation module:* In this module, if user wants to upload a file user needs to get key from private cloud.
3) *File Uploading module:* User can upload a file into the private cloud by using convergent key.
4) *Authorized Duplicate Check Sceme module:* The public cloud performs duplicate check[13] directly and tells the user if there is any duplicate. Public Cloud can store and retrieve file. De-duplication has a removing duplicate file[13]. Its will find out duplicate file

B. Used Algorithm

1) *Convergent Encryption Technique:* A user derives a convergent key from each original data copy and encrypts the data copy with the convergent key**.** The key generation algorithm that maps a data copy to a convergent key. The symmetric encryption

algorithm that takes both the convergent key and the data copy as inputs and then outputs a cipher text. The decryption algorithm that takes both the cipher text and the convergent key as inputs and then outputs the original data copy and the tag generation algorithm that maps the original data copy and outputs a tag.

C. Given Input and Expected Output As per the Design

1) Authentication
Input: Provide username and password to get permission for access.
Output: Became authenticated person to request and process the request.

2) Convergent key generation
Input: Key will be generated by private cloud.
Output: User can get the convergent key from private cloud.

3) Uploading file
Input: uploading any file to the public cloud by using convergent key.
Output: It will be stored in the public cloud.

4) Authorized duplicate check scheme
Input: User requesting some file.
Output: Authorized duplicate check scheme will check particular file is already existing or not

D. Application
1) CtrlS Real Cloud: The CtrlS Real Cloud has a multi-layered management model. The cloud controller server enables everything, from system architecture to VM root access, to be managed via the user interface and API. Real Cloud enables you to put up applications and manage them, all remotely and with utmost ease.
2) Cloud Layer Services: Discover the promise of cloud, not the compromises. Cloud Layer includes virtual servers, remote storage and a robust content delivery network that leverage our core advantages and longtime leadership in automated, on-demand, self-managed infrastructure.

III. PRACTICAL RESULT

The practical working for to verify the ownership de-duplication data is successful in cloud storage.
The screenshots are given below,

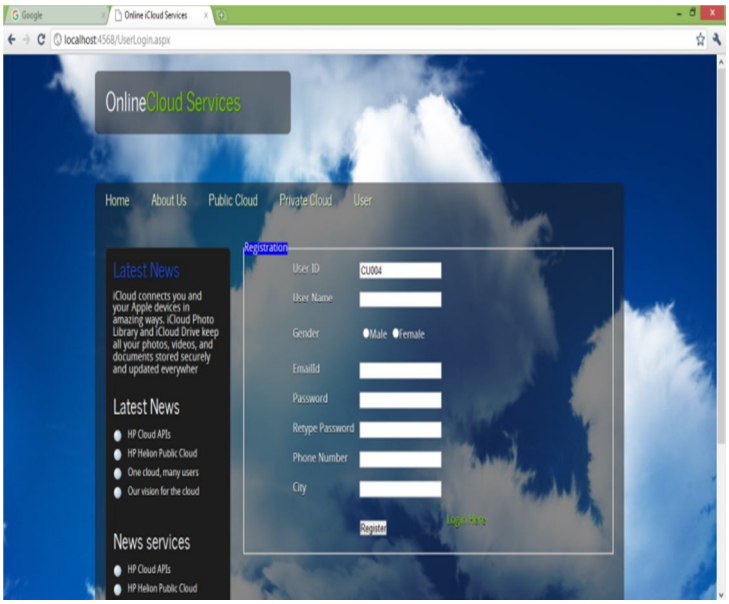


Figure 1: Registration page

Volume 4 Issue X, October 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

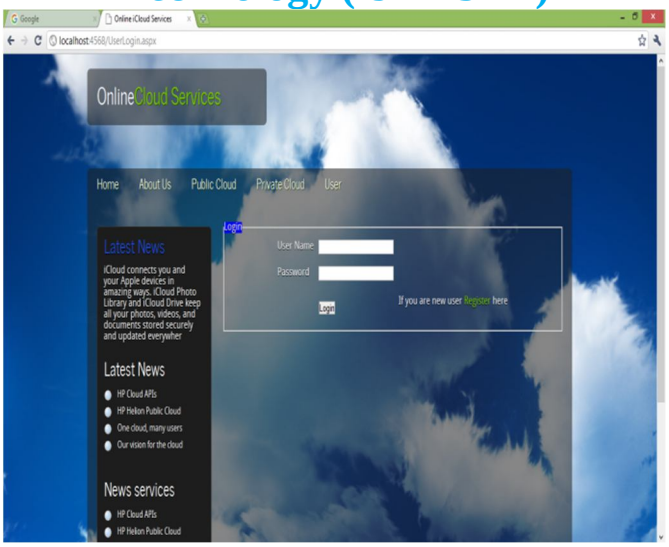


Figure 2: Login Page

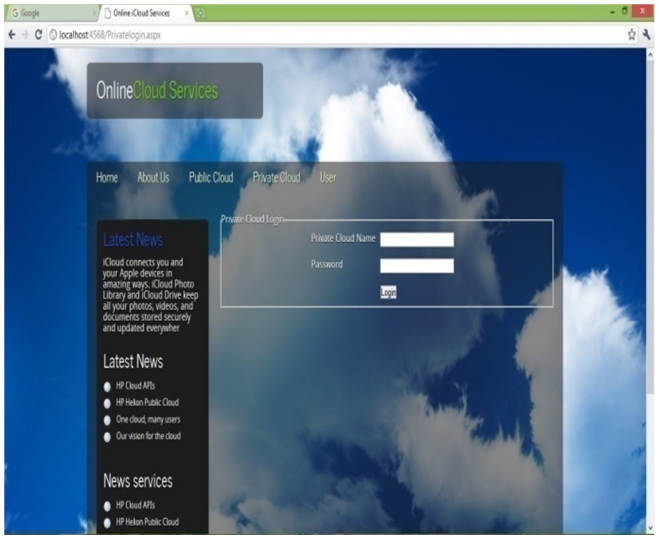


Figure 3 Private Cloud Login Page

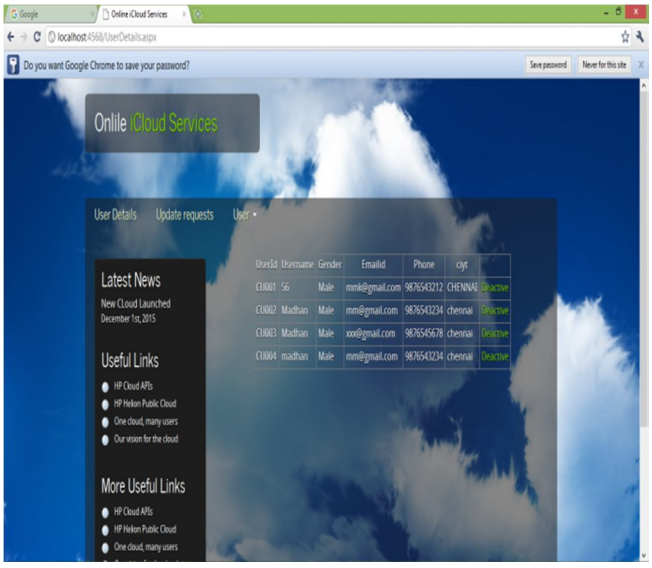


Figure 4 Activation page for users

Volume 4 Issue X, October 2016
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

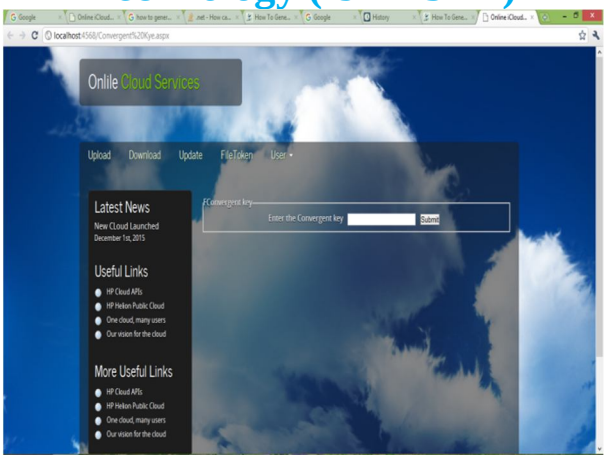


Figure 5 Convergent Key Generation page

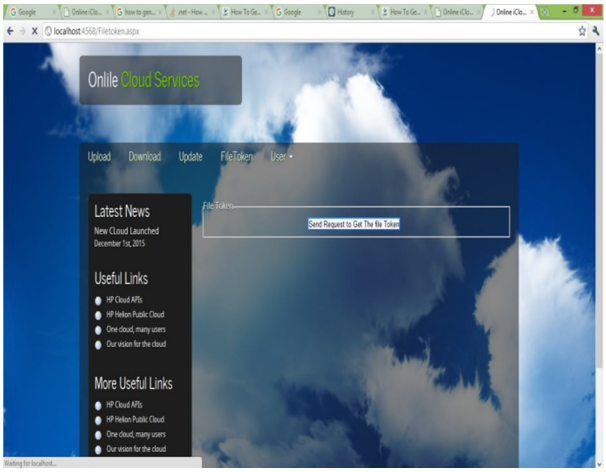


Figure 6 Send Request

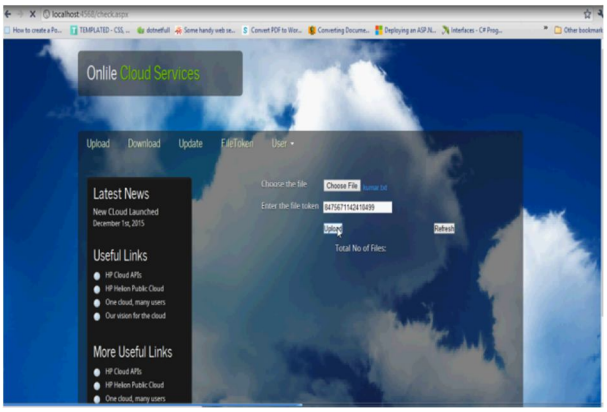


Figure 7 File uploading page

IV. FUTURE WORK

A. *Future Concept*

We design and implement a new system which could protect the security for predictable message. The main idea of our technique is that the novel encryption key generation algorithm. For simplicity, we will use the hash functions to define the tag generation functions and convergent keys in this section. In traditional convergent encryption, to support duplicate check, the key is derived from the file by using some cryptographic hash using some cryptographic hash function. To avoid the deterministic key generation, the encryption key for file in our system will be generated with the aid of the private key cloud server with privilege.

Future Technique: - Novel encryption key generation algorithm.

Technique Definition:-In dynamic environments, a new encrypted shared key has to be generated for every join/leave event and forwarded to the key distribution centre (KDC) of the requester. A novel Enhanced Encryption Algorithm (EEA) for generating a secured (encrypted) shared key is proposed for the transmission of packets in dynamic environments.

V. CONCLUSION

In this the notion of authorized data de-duplication was proposed to protect the data security by including differential privileges of users in the duplicate check. In which the duplicate-check tokens of files are generated by the private cloud server with private keys. The Scope of the project is eliminating duplicate copies of repeating data. Data de-duplication is one of important data compression techniques for eliminating duplicate copies of repeating data, and has been widely used in cloud storage to reduce the amount of storage space and save bandwidth. To protect the confidentiality of sensitive data while supporting de-duplication, the convergent encryption technique has been proposed to encrypt the data before outsourcing.

REFERENCES

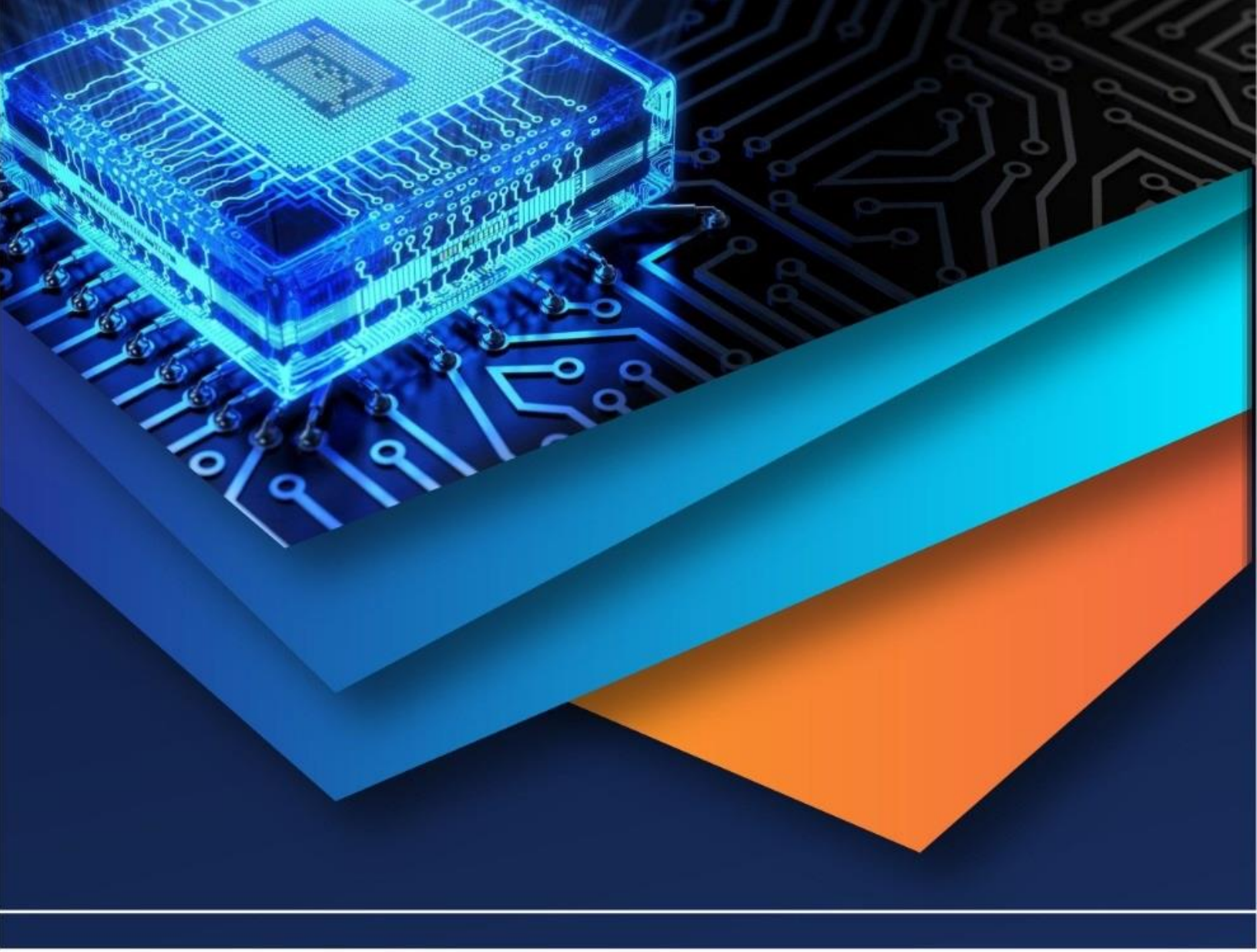
[1] OpenSSL Project, (1998). [Online]. Available: http://www.openssl.org/
[2] P. Anderson and L. Zhang, “Fast and secure laptop backups withencrypted de-duplication,” in Proc. 24th Int. Conf. Large InstallationSyst. Admin., 2010, pp. 29–40.
[3] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Dupless: Serveraidedencryption for deduplicated storage,” in Proc. 22nd USENIXConf. Sec. Symp., 2013, pp. 179–194.
[4] Google Drive : http://drive.google.com/. (Cited on page3
[5] M. Bellare, S. Keelveedhi, and T. Ristenpart, “Message-lockedencryption and secure deduplication,” in Proc. 32nd Annu. Int.Conf. Theory Appl. Cryptographic Techn., 2013, pp. 296–312.
[6] M. Bellare, C. Namprempre, and G. Neven, “Security proofs foridentity-based identification and signature schemes,” J. Cryptol.,vol. 22, no. 1, pp. 1–61, 2009.
[7] Dropbox,a file storaage ans sharing service.http://www.dropbox.com/.(cited on page 3.)
[8] Mark Lillibridge ,/kave Eshghi,Deepavali Bhagwat,Vinay Deolalikar,greg Trezise and Peter Camble”Sparse Indexing:Large Scale ,Inline Deduplication Using Sampling and Lacality”in HB Labs,UC Santa Cruz,HP Storae Works Division first.last @hb.com
[9] Pasquale puziosecludit and eurecom, refikmolvaeurecom, melek oneneurecom,”Clouded up : secure deduplication with encrypted data for cloud storage”
[10] M. Bellare and A. Palacio, “Gq and schnorr identification schemes:Proofs of security against impersonation under active and concurrentattacks,” in Proc. 22nd Annu. Int. Cryptol. Conf. Adv. Cryptol.,2002, pp. 162–177.
[11] N.O.Agrawal, Prof.Mr.S.S.Kulkri,”Secure deduplication and data security with efficient and reliable CEKM”
[12] S. Bugiel, S. Nurnberger, A. Sadeghi, and T. Schneider, “Twinclouds: An architecture for secure cloud computing,” in Proc.Workshop Cryptography Security Clouds, 2011, pp. 32–44.
[13] J. R. Douceur, A. Adya, W. J. Bolosky, D. Simon, and M.Theimer, “Reclaiming space from duplicate files in a serverless distributed file system,” in Proc. Int. Conf. Distrib. Comput. Syst., 2002, pp. 617–624.
[14] D. Ferraiolo and R. Kuhn, “Role-based access controls, ” in Proc.15th NIST-NCSC Nat. Comput. Security Conf., 1992, pp. 554–563.
[15] JadapalliNandhini, RamireddyNavateja Reddy, “Implementation of Hybrid Cloud Approach for Secure Authorized Deduplication”, International Research Journal of Engineering and Technology, vol 2, issue 3, jun 2015.
[16] Gaurav Kakariyaand SonaliRangdale, “A Hybrid Cloud Approach For Secure Authorized Deduplication”, International Journal of Computer Engineering and Applications, Volume 8, Issue 1, oct 2014.
[17] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraided encryption for deduplicated storage. In USENIX Security Symposium, 2013.
[18] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl, “A secure data deduplication scheme for cloud storage,” Tech. Rep. IBM Research, Zurich, ZUR 1308-022, 2013.
[19] J. Li, X. Chen, M. Li, J. Li, P. Lee, andW. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.
[20] Jin Li, Yan Kit Li, Xiaofeng Chen, Patrick P. C. Lee, Wenjing Lou, “A Hybrid Cloud Approach for Secure Authorized Deduplication”, IEEE Transactions on Parallel and Distributed Systems, 2014.
[21] Salvatore J. Stolfo, Malek Ben Salem and Angelos D. Keromytis “Fog Computing: Mitigating Insider Data Theft Attacks in the Cloud” IEEE Symposium On Security And Privacy Workshop (SPW) YEAR 2012.
[22] N.O.Agrawal, S.S.Kulkarni”Secure Deduplication and Data Security with efficient and reliable CEKM” IJAIEM Transition On parallel And Distributed System,VOL.3, Issue. 11, November 2014.
[23] J. Yuan and S. Yu, “Secure and constant cost public cloud storage auditing with deduplication,” IACR Cryptology ePrint Archive, 2013:149, 2013.
[24] J. Li, X. Chen, M. Li, J. Li, P. Lee, and W. Lou. Secure deduplication with efficient and reliable convergent key management. In IEEE Transactions on Parallel and Distributed Systems, 2013.

AUTHOR DETAILS

Sreedarsa K S, B-Tech(CES), ME (CSE)’, is pursuing Master of Engineering from Maharajas Institute of Technology. She passed BTech in the year 2007 and worked as a guest lecturer for 5 years and then as Management Information System coordinator around 4 months.

Prof. T. Sivakumar, ME (CSE), He had 10 years of experience in teaching industry. His areas of interest are data mining and data source.Participated in various international and national level conference, seminars/webinars and workshops.

Neethu Rosiline, B.E(CSE)., M.E(CSE)', is pursing Master of Engineering from Maharajas Institute of Technology. She passed BE in the year 2007 and worked in the IT industry for more than 2 years and then worked as a guest lecturer for about 1 year.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY