International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 06
May 2017

Secure and Efficient Deduplication Using Hybrid Cloud Approach

AL-GARA'AWI AHMED HADI HUSSEIN

Master Student at faculty of Engineering in Foreign Languages (FILS) / University of Politehnica, Romania/Bucharest

Computer engineer in IT and communication department at Education directorate in Diwaniyah/Iraq

E-mail: ahmed.id.hadi@gmail.com

ABSTRACT

In cloud computing, the big attention always is on security, about the abilities in which the shared data must be stored in high level security system, not all the users can access in the cloud, just who has the privilege given by the cloud itself or by who owns the data and so on. In addition, and as we all know that the customers are buying clouds to get benefits of never mind about the infrastructure and to be sure that all their data will totally be in safe. There is also big attention on how to never duplicate the files in cloud in order to reduce the amount of storage space because as we know that the customer is paying to get the storage and for sure nobody wants his data to be repeated in his system. For that, this paper will discuss a type of deduplication of files securely by giving just the owner the rights to choose the users who will be able to get access and have the right to upload or download files, and the use of AES algorithm as a way to encrypt/decrypt.

Key words: cloud computing, deduplication, encryption, decryption, AES, private cloud, public cloud, cipher text

INTRODUCTION

Cloud computing give us unlimited services which cross the whole Internet, while is hiding the detail of platform and implementation .The providers of cloud services give us in the same time highly available storage and parallel computing resources at low cost. Because cloud computing is becoming prevalent, amount of data is increasing and it is stored in the cloud and can be shared by the users with specified privileges and defining the access rights of the stored data. For storage services it is a big challenge in cloud computing to manage the ever-increasing volume of data. In cloud computing we need deduplication for storage services and for making data management scalable, which is a good technique. The specialized data compression for deleting large copies of replicated storages data is named data deduplication. The procedure is used to improve usage of storage and can be applied to transfers of data to lower the number of bytes sent.

Data deduplication erase unnecessary data by keeping only one physical copy and sending another unnecessary data to that copy. This can occur at each level of the file or the block level. For file level deduplication, it erase the duplicate copies of the same file. Also deduplication can be at the same place at the block level, which erase duplicate blocks of data that appear in different files. Data deduplication has a lot of advantages, but also appear issues in the security and privacy of users protected data

that are affected to attacks. The traditional encryption is offering confidentiality of data and is different from data deduplication. But traditional encryption need non-identical users to encode their data with their own keys. So, the similar copies of various data users will generate non-identical codes, for that, the deduplication will not be available. And for making possible the deduplication, encryption become available by enforcing the privacy of data.

For the encryption, is used a convergent key, which become available by calculating the cryptographic hash value of the data copy. After generating the key and encryption of data, users keeps the keys and send the secret code text to the cloud. Because the operation of encryption is available and it is generated from the content of data, the same data of copies will obtain the convergent key and the same secret code text. For preventing unapproved access is required a secure of property protocol and for generating the proof that the user really holds the same file when a copy is found. After the proof, the next users with the same file will receive a pointer from the server and it will not be necessary to send the same file.

A user can send the encrypted file with the pointer from the server, which can only be decrypted by the identical data owners with their keys. So, the AES encryption is giving the cloud the permission to access deduplication on the cipher texts and the proof of ownership prevents the unapproved user to use the file. With all of this, the last deduplication of systems cannot handle to differential authorization duplicate check, which is necessary in various applications. In such an authorized deduplication system, each user is receiving a set of privileges during system initialization. Every document sent to the cloud is also limited by a set of privileges to specify which kind of users is allowed to perform the double check and access the documents. For sending his duplicate check solicitation for some document, the user must take it and also his privileges as inputs. The user can obtain a duplicate for the document if there is a copy of this and a matched privilege stored in cloud. For example, in a company, many different privileges will be assigned to employees. In order to save cost and efficiently management, the data will be moved to the storage server provider (SCSP) in the public cloud with specified privileges and the deduplication technique will be applied to store only one copy of the same file. To have security, some files will be encrypted and to allow duplicate verification by employees with stated privileges to use control of access. We can say, that in the deduplication based on convergent encryption, no differential privileges which have been considered technique. It seems to be opposite if we want to realize both deduplication and differential authorization duplicate check at the same time.

EXISTING SYSTEM:

- Data deduplication systems, the non-public cloud is concerned as a proxy to permit information owner/users to firmly perform duplicate visit differential privileges.
- Such design is sensible and has attracted abundant attention from researchers.
- The information homeowners solely source their information storage by utilizing public cloud whereas the info operation is managed privately cloud.

DISADVANTAGES OF EXISTING SYSTEM:

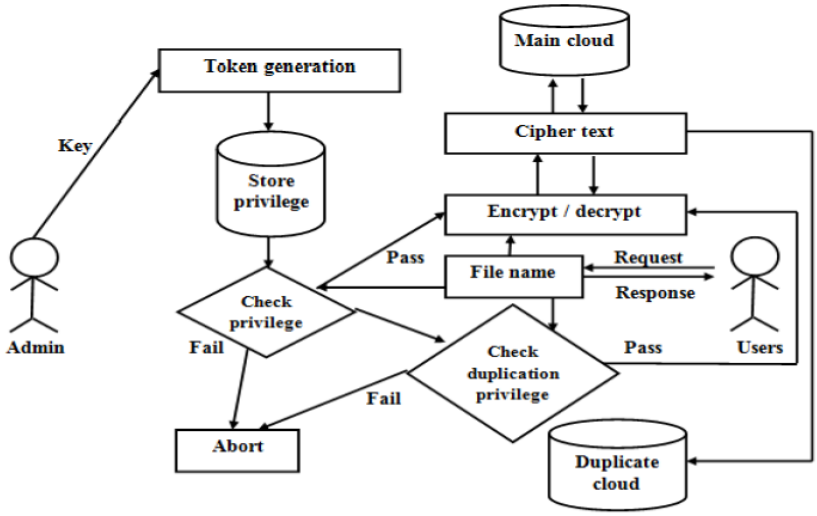
- Traditional encryption, while providing data confidentiality, is incompatible with data deduplication.
- Identical data copies of different users will lead to different cipher texts, making deduplication impossible.

PROPOSED SYSTEM:

In this paper, we tend to enhance our system in security. Specifically, we tend to gift a complicated theme to support stronger security by encrypting the file with differential privilege keys. During this manner, the users while not corresponding privileges cannot perform the duplicate check. Moreover, such unauthorized users cannot rewrite the cipher text even conspire with the S-CSP. Security analysis demonstrates that our system is secure in terms of the definitions laid out in the planned security model.

ADVANTAGES OF PROPOSED SYSTEM:

- The user is only allowed to perform the duplicate check for files marked with the corresponding privileges.
- We present an advanced scheme to support stronger security by encrypting the file with differential privilege keys.
- Reduce the storage size of the tags for integrity check. To enhance the security of deduplication and protect the data confidentiality,



System architecture [5]

Why Hybrid cloud

A hybrid cloud is an integrated cloud service that uses both private and public cloud systems to perform distinct functions within the same organization.

In general, all cloud services must provide a certain level of efficiency in response to different levels of use. From this point of view, public cloud solutions are likely to be more cost-effective and scalable than private cloud. Therefore, an organization can

maximize efficiency by using integrated public cloud solutions for all current operations, and rely on private cloud solutions only in cases which requires a higher degree of security

Hybrid cloud models can be implemented in several ways:

- Different cloud providers associate and provide both public and private cloud services as an integrated service - Individual cloud service providers offer full cloud hybrid packages

- Some businesses and organizations manage cloud privately, but also cloud a public cloud service that they later integrate into their own infrastructure

Some practical examples of using hybrid cloud modules:

1. A company that owns an online store The online store is hosted using a private cloud - the safest and most scalable solution, but the presentation brochures and catalogs are hosted in a public cloud - the most cost effective solution (in this case, the main concern is the cost, not the level of security).

2. An IaaS service (Infrastructure as a Service) offered to a financial firm Private cloud for client and public cloud database storage for documents used in collaborative projects, documents that can be accessed by multiple users from any convenient location.

A hybrid cloud configuration such as hybrid hosting can offer users the following benefits:

- Scalability

While the private cloud offers a certain level of scalability depending on its configurations (whether hosted internally or externally), public cloud will offer less scalability as it generally has the resources of a cloud infrastructure big. By moving as many common security-less common functions as possible to the cloud audience, the benefits of a high level of scalability while reducing private cloud requirements

- Cost Efficiency

According to the same principle, public cloud solutions are likely to offer more significant savings in scalability (such as centralized management), being more cost-effective than private cloud solutions. Therefore, hybrid cloud systems allow their users to benefit from these savings for all less sensitive features that do not require a high degree of private cloud security.

- Security

The private cloud element of the hybrid cloud model does not only provide the desired security level when required, but may also satisfy any regulations regarding the handling and storage of data where it is applicable.

- Flexibility The availability of both types of resources: private cloud and scalable public cloud can provide organizations with more opportunities to explore new ways to get operational revenue

1.2 Advantages of using hybrid cloud

An important benefit is the ownership of a local private infrastructure that is directly accessible, not via the Internet. This reduces the latency compared to public cloud services and the unique blocking point that the internet may be when it does not work disappears.

Another benefit is the ability to have an infrastructure that supports the company's average needs, while retaining the ability to use the private cloud for circumstances where processing needs outweigh the

available cloud power. The costs incurred by a company that has a few annual demand peaks are lower when these peaks are solved by the public cloud than when a more powerful local infrastructure is in place that is partially unused almost all year.[9]

1.3 Benefits of data deduplication

- Storage-based data deduplication reduces the amount of storage needed for a given set of files. It is most effective in applications where many copies of very similar or even identical data are stored on a single disk. In the case of data backups, which routinely are performed to protect against data loss, most data in a given backup remain unchanged from the previous backup. Common backup systems try to exploit this by omitting files that haven't changed or storing differences between files. Neither approach captures all redundancies, however. Hard-linking does not help with large files that have only changed in small ways, such as an email database; differences only find redundancies in adjacent versions of a single file (consider a section that was deleted and later added in again, or a logo image included in many documents).[4]
- Network data deduplication is used to reduce the number of bytes that must be transferred between endpoints, which can reduce the amount of bandwidth required. [2]
- Virtual servers and virtual desktops benefit from deduplication because it allows nominally separate system files for each virtual machine to be coalesced into a single storage space. At the same time, if a given virtual machine customizes a file, deduplication will not change the files on the other virtual machines—something that alternatives like hard links or shared disks do not offer. Backing up or making duplicate copies of virtual environments is similarly improved.[4][5]

AES encryption/decryption

AES (Advanced Encryption Standard), also known as Rijndael, is a standardized algorithm for symmetric block encryption, widely used today in applications.

AES is based on a design principle known as a substitution-permutation network, a combination of both substitution and permutation, and is fast in both software and hardware. [11] AES has a fixed block size of 128 bits, and a key size of 128, 192, or 256 bits. By contrast, the Rijndael specification is specified with block and key sizes that may be any multiple of 32 bits, both with a minimum of 128 and a maximum of 256 bits.

AES operates on a 4 × 4 column-major order matrix of bytes, termed the state, although some versions of Rijndael have a larger block size and have additional columns in the state. Most AES calculations are done in a particular finite field. [10]

For instance, if there are 16 bytes, b1, b2, b3… b15, these bytes are represented as this matrix:

The key size used for an AES cipher specifies the number of repetitions of transformation rounds that convert the input, called the plaintext, into the final output, called the ciphertext. The number of cycles of repetition are as follows:

- 10 cycles of repetition for 128-bit keys (used in our system).

- 12 cycles of repetition for 192-bit keys.
- 14 cycles of repetition for 256-bit keys.

Each round consists of several processing steps, each containing four similar but different stages, including one that depends on the encryption key itself. A set of reverse rounds are applied to transform ciphertext back into the original plaintext using the same encryption key. [12]

3.1Description of algorithm [11]

1. Key Expansions: round keys are derived from the cipher key using Rijndael's key schedule. AES requires a separate 128-bit round key block for each round plus one more.
2. Initial Round
    1. AddRoundKey—each byte of the state is combined with a block of the round key using bitwise xor.
3. Rounds
    1. SubBytes—a non-linear substitution step where each byte is replaced with another according to a lookup table.
    2. ShiftRows—a transposition step where the last three rows of the state are shifted cyclically a certain number of steps.
    3. MixColumns—a mixing operation which operates on the columns of the state, combining the four bytes in each column.
    4. AddRoundKey
4. Final Round (no MixColumns)
    1. SubBytes
    2. ShiftRows
    3. AddRoundKey.

There are three entities define in our system as shown in figure 1, those are,

1. Users
2. Private cloud which represents the store of privileges in our system.
3. S-CSP which represents the public cloud

De-duplication performed by S-CSP by checking if the contents of two files are the same and stores only one of them. [9][19]

- ❖ S-CSP (public cloud). This is an entity that provides a data storage service in public cloud. The S-CSP provides the data outsourcing service and stores data on behalf of the users. To reduce the storage cost, the S-CSP eliminates the storage of redundant data via de-duplication and keeps only unique data. In this paper, we assume that S-CSP is always online and has abundant storage capacity and computation power.

- ❖ Data Users. A user is an entity that wants to outsource data storage to the S-CSP and access the data later. In a storage system supporting de-duplication, the user only uploads unique data but does not upload any duplicate data to save the upload bandwidth, which may be owned by the same user or different users. In the authorized de-duplication system, each user is issued a set of privileges in the setup of the system. Each file is protected with the convergent encryption key and privilege keys to realize the authorized de-duplication with differential privileges.

- Private Cloud. Compared with the traditional de-duplication architecture in cloud computing, this is a new entity introduced for facilitating user's secure usage of cloud service. Specifically, since the computing resources at data user/owner side are restricted and the public cloud is not fully trusted in practice, private cloud is able to provide data user/owner with an execution environment and infrastructure working as an interface between user and the public cloud. The private keys for the privileges are managed by the private cloud, who answers the file token requests from the users. The interface offered by the private cloud allows user to submit files and queries to be securely stored and computed respectively.

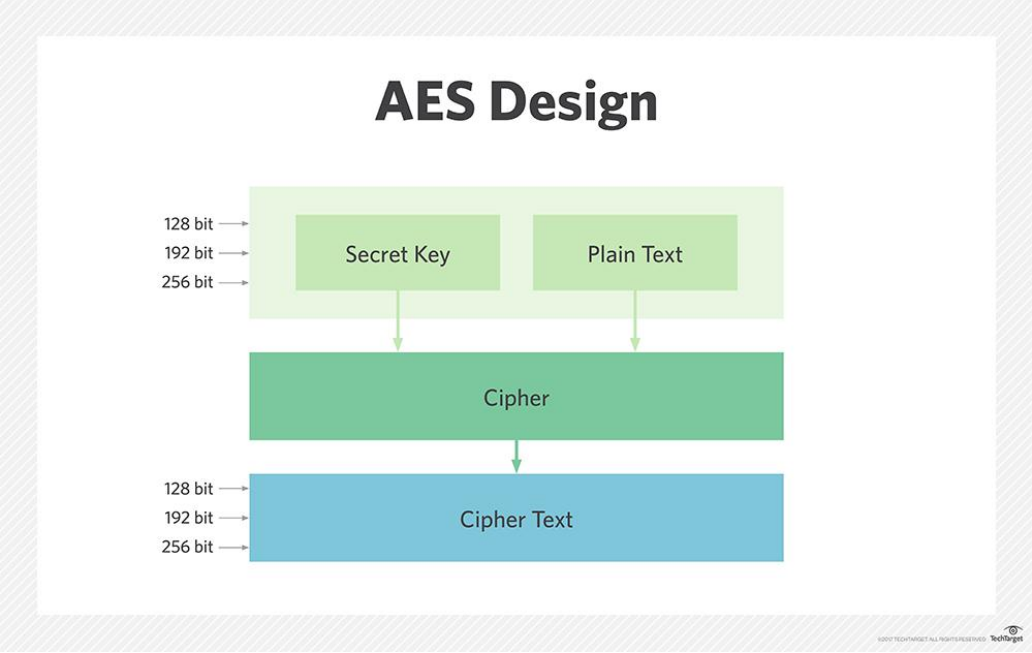
Hybrid clouds generally having twin clouds (private cloud and public cloud). This architecture is used for data de-duplication. For example, an enterprise might use a public cloud service, such as Amazon S3, for archived data, but continue to maintain in-house storage for operational customer data. Alternatively, the trusted private cloud could be a cluster of virtualized cryptographic co-processors, which are offered as a service by a third party and provide the necessary hardware based security features to implement a remote execution environment trusted by the users.[3]

4.3Encryption and decryption by AES

Advanced Encryption Standard is a symmetric block cipher to protect classified information and is implemented in software and hardware throughout the world to encrypt sensitive data.

How AES works?

AES comprises three block ciphers: AES-128, AES-192 and AES-256. Each cipher encrypts and decrypts data in blocks of 128 bits using cryptographic keys of 128-, 192- and 256-bits, respectively. [11]



SYSTEM Components:

- Private cloud

Private cloud is the component responsible for storing privileges, which include encrypted keys, and handling deduplication. Private cloud contains a small database in order to keep track of file ownerships, file composition and avoid the data to be duplicated. The tables used for this purpose are file, pointer and signature tables.

The tables used by private cloud are:
_ Client table. Which contains

NAME	USERNAME	PASSWORD	MAIL	STATUS	ACTION

_ Client request table. Which contains:

USERNAME	REQUEST	TIME	ACCEPT	DENY

When users ask to get access a file, private cloud will checks if the user who did the request is authorized to get that file. By this way, private cloud will ensure that this user is not accessing the data of someone else. This operation can be considered as an additional access control mechanism.

- Public cloud

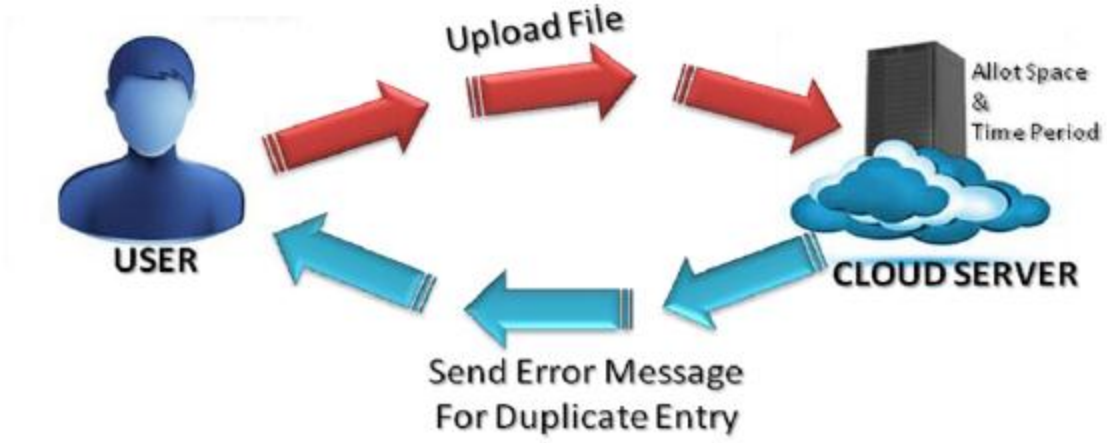
  It has three main parts:
  - authenticating users during the storage/retrieval request,
  - performing access control by verifying block signatures embedded in the data,
  - Encrypting/decrypting data traveling from users to the cloud and vice versa.

The server takes care of adding an additional layer of encryption to the data uploaded by users. Before being forwarded to private cloud, data are further encrypted in order to prevent private cloud and any other component from performing dictionary attacks and exploiting the well-known weaknesses of AES encryption. During file retrieval, blocks are decrypted and the server verifies each block with the user's public key. If the verification process fails, blocks are not delivered to the requesting user.

The tables used in public cloud:

- Uploads which includes file name, owner name, upload time and the size of the uploaded file.

FILE NAME	OWNER NAME	UPLOAD TIME	SIZE



Uploading file process [12]

- Downloads which includes file name for the downloaded file, user name and download time

FILE NAME	USER NAME	DOWNLOADED TIME

- Updates in which there are file name, user name and downloaded file.

FILE NAME	USER NAME	DOWNLOADED TIME

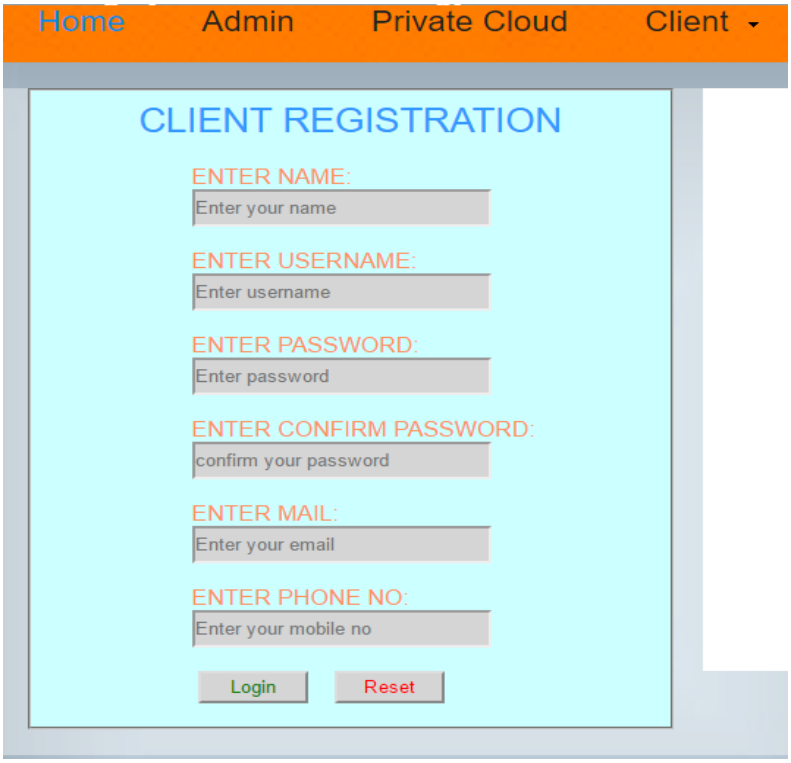
We can say that the public cloud represents the data owner which is related with all operations during the system work

User

The part of the user is going to get the right to get access to be able to upload or to download files, by passing the AES encryption/decryption mechanism. In addition, the user also encrypts each key derived from the corresponding block with the previous one and his secret key in order to outsource the keying material as well and thus only store the key derived from the first block and the file identifier. For each file, this key will be used to decrypt and re-build the file when it will be retrieved. Instead, the file identifier is necessary to univocally identify a file over the whole system.

It has two parts: registration and sign in

In registration part we have name, user name, password, and e-mail and phone number



IMPLEMENATION

Deployment of secure authorized deduplication System in private cloud

This paper focuses on a deployment models such as platform as a service, software as a service and Infrastructure as a service using UEC. A Private Server program is used to model the private cloud which manages the private key and handles the file token computation. A Storage Server program is used to model the SCSP which stores and de duplicates files. The hybrid cloud approaches for securing the deduplication is developed using J2EE, a user can access from any computer connected to the private cloud by using Apache web server.

5.1 OUTPUT SCREENS:

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective. The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

Data Owner: - Data owner will make account in our application by using the registration form and by using the his/her user name and password he can login in to our application they can upload and download data from our cloud server the data will be provide security by encrypting the data in the files and a key will be generated for every file that upload in the cloud data base.

Users: - Users will make account in our application by using the registration form and by using the his/her user name and

password he can login in to our application they download data from our cloud server the data by using the token key provided by the data owner and decrypt the data also

Usage is the phase of the undertaking when the hypothetical configuration is transformed out into a working system. Accordingly it can be thought to be the most basic stage in accomplishing a fruitful new system and in giving the client, certainty that the new system will work and be compelling.

The execution stage includes cautious arranging, examination of the current system and its limitations on usage, planning of techniques to accomplish changeover and assessment of changeover strategies.

MODULE DESCRIPTION:

File Confidentiality:

The outline objective of record classification requires to keep the cloud servers from getting to the substance of documents. Uncommonly, we require that the objective of document classification should be impervious to "word reference assault". That is, even the foes have pre-information of the "lexicon" which incorporates all the conceivable documents, regardless they can't recoup the objective record

Secure Deduplication:

Deduplication is a method where the server stores just a solitary duplicate of every document, paying little respect to what number of customers requested that store that record, such that the plate space of cloud servers and in addition system transfer speed are spared. Be that as it may, insignificant customer side deduplication prompts the spillage of side channel data. For instance, a server telling a customer that it need not send the document uncovers that some other customer has precisely the same, which could be delicate data for some situation.

Encryption & Decryption:

Encryption and unscrambling gives information secrecy in deduplication. A client (or information proprietor) gets a united key from the information content and scrambles the information duplicate with the focalized key. Moreover, the client determines a tag for the information duplicate, such that the tag will be utilized to distinguish copies. Here, we expect that the label accuracy property holds, i.e., if two information duplicates are the same, then their labels are the same. Formally, a merged encryption plan can be characterized with four primitive capacities:

Integrity Auditing:

The primary configuration objective of this work is to give the capacity of checking rightness of the remotely put away information. The respectability check further requires two elements:

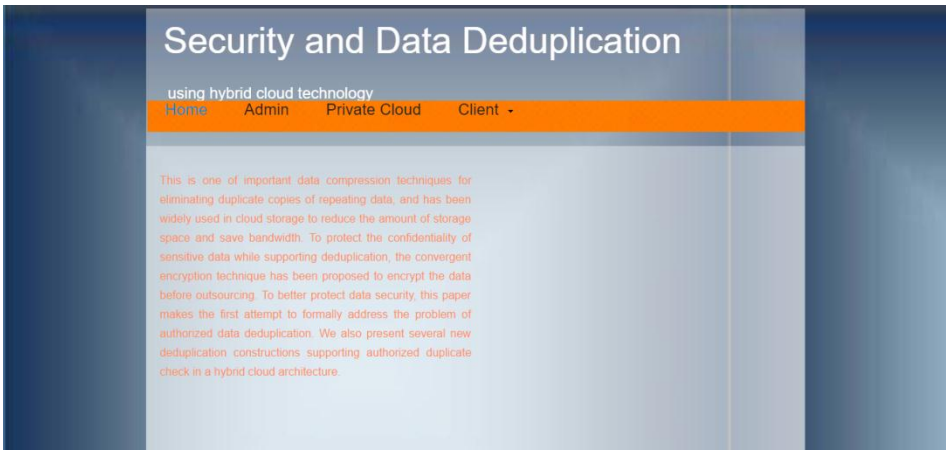
1. Public check, which permits anybody, not only the customers initially put away the record, to perform confirmation;

2. Stateless check, which can kill the requirement for state data upkeep at the verifier side between the activities of inspecting and information stockpiling.

Steps:-

Home page

It represents the main page of the system, in which will find all the main components those are we used them, Public cloud, Private cloud and Clients which involved sign in and registration. And short description about the system.



1. Private Cloud

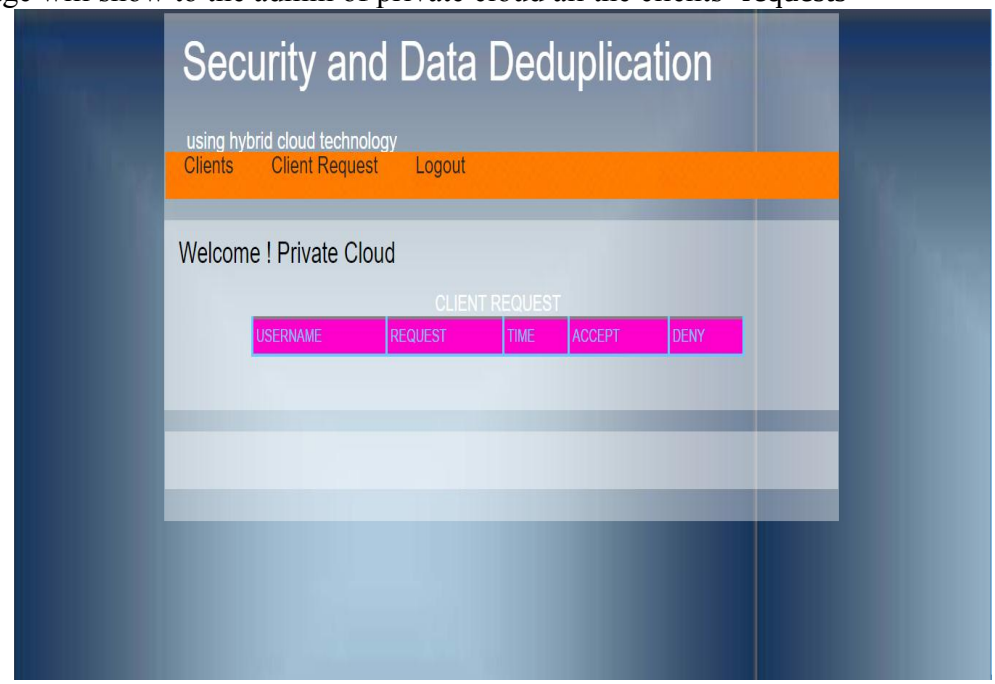
The private before getting access by the admin will be:

After the admin will get access by his user name and the password which he got by the system which stored in the database specified for that.
We are using the username and the password: cloud

We can see that the private cloud including everything about the users and the privileges which had been gotten for them by the system

Clients represent the information which the clients used used them to register into the system, these are NAME, USERNAME, PASSWORD, E_MAIL, STATUS and ACTION.

Client request represents which actions the client wants to get, the upload or the download, and this page will show to the admin of private cloud all the clients' requests

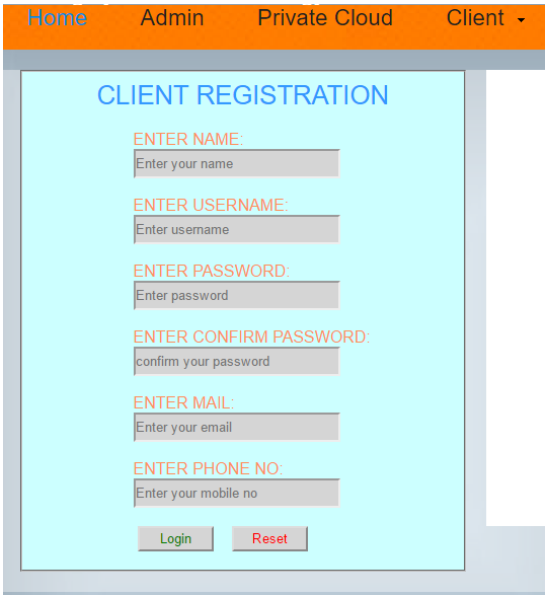


2. Clients

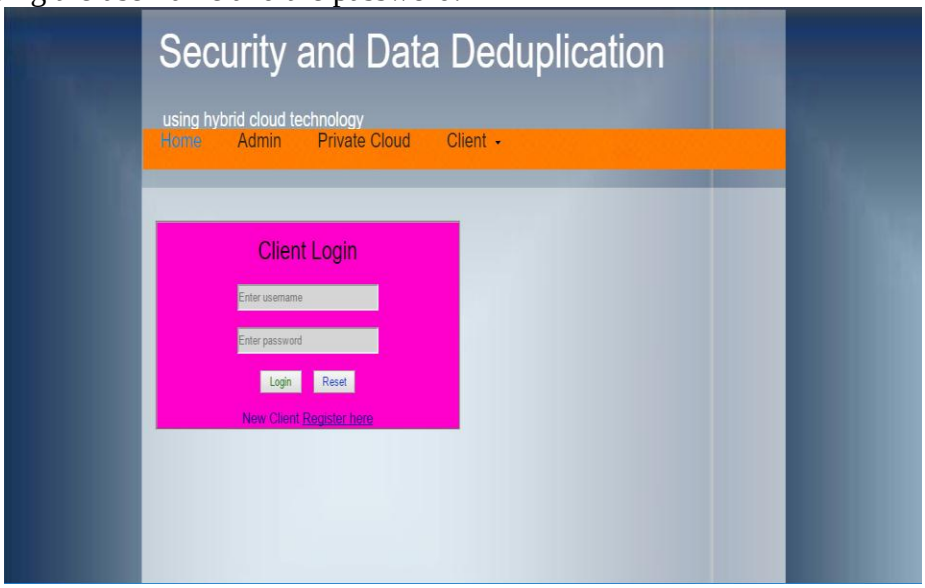
This represents two parts: sign in and registration.

For the registration, just the new client will use it to, if he really wants to get access and to get the right to upload or download the required files.

It includes Name, Username, and password, confirm password, e-mail and Phone number. All of them is chosen by the user and he has the right to choose what he like, but he has to use his real e-mail because he will receive the token code which he will use it to activate his account, otherwise, the client will not be able pass the registration process.



After the client will receive his token code, he will use it to submit to be able to sign into the system by using the username and the password.



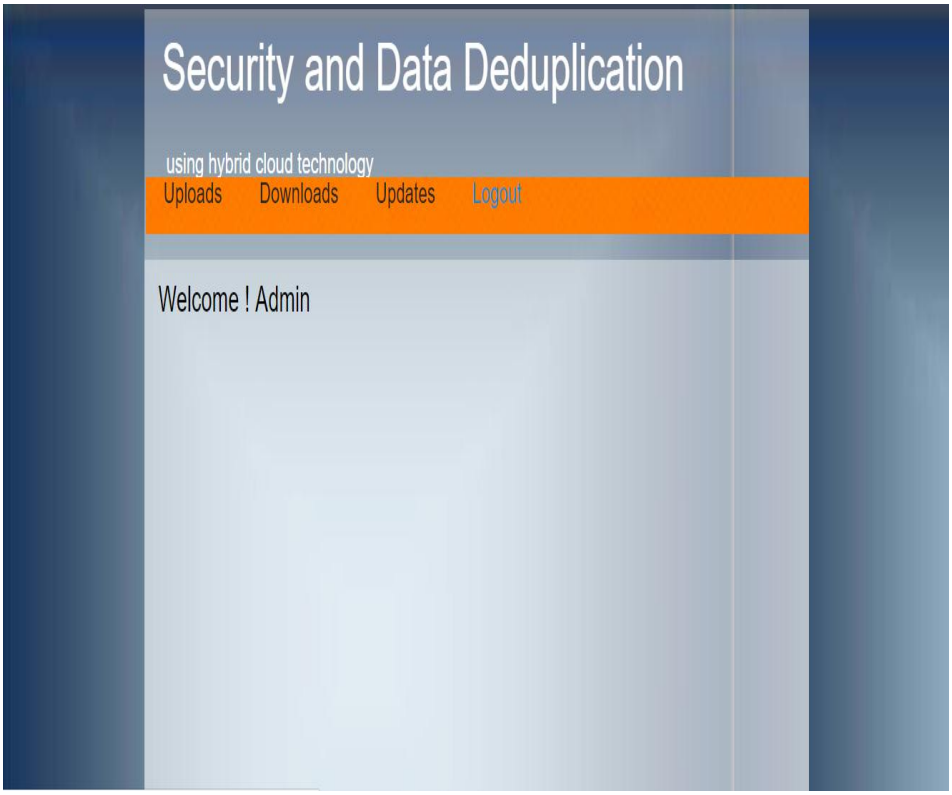
When the client submit and get the access, he will ask the system to activate the action that he wants (uploading or downloading or the both), all these activities are controlled by the private cloud. After the private cloud gives the agreement to the client to use the system, the client will get access to the public cloud in which he will find where to upload or to download the files.

3. Public cloud

In public cloud, all the users who got the right to access the system by the private cloud will be able to access it. The admin we are using represented by the username and

password: admin. Just who has the right from private cloud as an admin in public cloud will be able to access here.

The admin of public cloud will has the control of the files in this system, it represents the main goal of this, the ability of sharing the files.



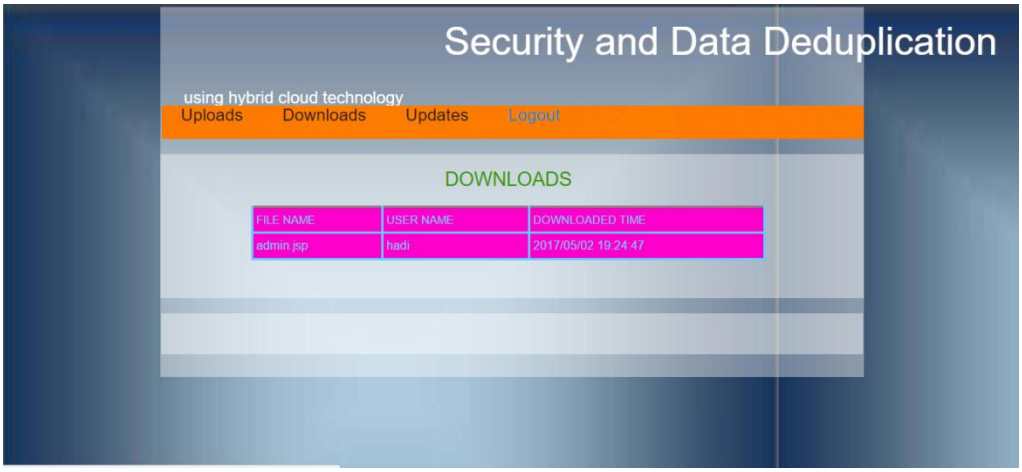
We can see in public cloud that we have uploads, downloads and updates. In Uploads, it's so necessary to put file name, owner name, upload time and the size, it will be used into the mechanism of deduplication by checking if the file wanted to be uploaded are there in this table

or no, if it's found, the uploading process will be denied, but if it's not found, the client will upload it successfully.

In uploading process, we can say that the process will happen in these steps:

Start
Step 1: Read file, the file that the client wants to upload.
Step 2: cloud server checks for duplication
Step 3: sends duplication response whether the file already exists or not.
Step 4: if the file does not exist
Print "file does not exist".
Step 5: then it uploads the file.
Step 6: if the file already exist
Print "file already exist".
Stop.

In downloading process, will have file name, username for the user who downloaded it and the downloading time, and here any user (who got the right to access) can download any file even many times.



This process will happen in these steps:

Start
Step 1: read file
Step 2: cloud server checks the file if it's still exist not
Step 3: if the file exist
Print "file exist"
Step 5: then it downloads the file
Step 6: if the file does not exist
Print "file does not exist"
Stop

CONCLUSION

In this case for protecting the data security it was used the data deduplication by including different benefits of users in the duplicate check. For checking authorized duplicate in hybrid cloud architecture, we used many new deduplication constructions, in which the duplicate-check tokens of files are created by the private cloud server with private keys. Security analysis indicate that the schemes are safe in terms of insider and outsider attacks specified in the proposed security mode. Privileges of users in the duplicate check: Data deduplication technique is proposed .Data deduplication is lowering the storage overhead created by the big quantity of data because the duplicate copies are not stored in cloud. The model is using two clouds for addressing fault tolerance. It is used EAS because overcomes the drawbacks of traditional encryption. The paper was implemented to provide security as well as to reduce the storage overhead. Many tested experiments were performed for different parameters to check the efficiency of the system.

FUTURE WORK ABILITY

As part of future work, deduplication may be extended with more security features such as proofs of retrievability [34], data integrity checking [33] and search over encrypted data [13]. In this paper we mainly focused on the definition of the two most important operations in cloud storage that are storage and retrieval. We plan to define other typical operations such as edit and delete. After implementing a prototype of the system, we aim to provide a full

performance analysis. Furthermore, we will work on finding possible optimizations in terms of bandwidth, storage space and computation.And in our system the admins who are controlling, which means that the users will face the delay, in the future it will be perfect to make the managing can be done automatically after accepting all the information that the user registered.

RESOURCES

[1] Sunita S. Velapure, S. S. Barde, A Hybrid Cloud Approach for Secure Authorized Deduplication

[2] M.B. Benjula Anbu Malar, M. Lawanya Shri, M. Deepa and K. Santhi, Approach for Secure Authorized Deduplication using Hybrid Cloud

[3] JAdapalli Nandini , Ramireddy Navateja Reddy, Implementation Of Hybrid Cloud Approach For Secure Authorized Deduplication

[4] Pasquale Puzio, Refik Molva, Melek O¨ nen, Sergio Loureiro, ClouDedup: Secure Deduplication with Encrypted Data for Cloud Storage

[5] Srushty V.Gurav , Roopalakshmi.S, K. Narasimha Murthy, Authorized Deduplication System with Differential Privilege User Checks In Hybrid, May 2015.
Cloud

[6] https://www.cloudcomputing-news.net/news/2015/sep/03/why-hybrid-cloud-so-important-market-prediction-large/

[7] "In-line or post-process de-duplication? . Backup Central. Archived from the original on 2009-12-06.

[8] "Inline vs. post-processing deduplication appliances". Search data backup. Techtarget. com. 2009-10-16.

[9] www.koding.ro/software/solutii-cloud-productivitate-si-infrastructura/cloud-privat-si-hibrid/

[10] searchsecurity.techtarget.com/definition /Advanced-Encryption-Standard

[11] Adam Berent, ABT software development, Advanced Encryption Standard by Example

[12] Raj Jain, Advanced Encryption Standard(AES), 2011

[13] Mihir Bellare, Alexandra Boldyreva, and Adam ONeill. Deterministic and efficiently searchable encryption. In Advances in Cryptology-CRYPTO 2007, pages 535–552. Springer, 2007.

[14] Nicolae Goga, software testing, Faculty of Engineering in Foreign Languages, university of Politehnica

[15] Lu Luo, Software Testing Techniques Technology Maturation and Research Strategy, USA

[16] Adtha Lawanna, The Theory of Software Testing, Department of Information Technology, Faculty of Science and Technology Assumption University, Bangkok, Thailand, 2012

[17] 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.

[18] J. Yuan and S. Yu. Secure and constant cost public cloud storage auditing with deduplication.

[19] J. Xu, E.-C. Chang, and J. Zhou. Weak leakage-resilient client-side deduplication of encrypted data in cloud storage. In ASIACCS, pages 195–206, 2013.

[20] J. Stanek, A. Sorniotti, E. Androulaki, and L. Kencl. A secure data deduplication scheme for cloud storage. In Technical Report, 2013.

[21] A. Rahumed, H. C. H. Chen, Y. Tang, P. P. C. Lee, and J. C. S. Lui. A secure cloud backup system with assured

deletion and version control. In 3rd International Workshop on Security in Cloud Computing, 2011.

[22] R. D. Pietro and A. Sorniotti. Boosting efficiency and security in proof of ownership for deduplication. In H. Y. Youm and Y. Won, editors, ACM Symposium on Information, Computer and Communications Security, pages 81–82. ACM, 2012.

[23] C. Ng and P. Lee. Revdedup: A reverse deduplication storage system optimized for reads to latest backups. In Proc. of APSYS, Apr 2013.

[24] http://www.math.uni-hamburg.de/doc/java/tutorial/getStarted/intro/definition.html

[25] https://www.ntu.edu.sg/home/ehchua/programming/howto/Tomcat_HowTo.html

[26] https://docs.oracle.com/javase/tutorial/jdbc/

[27] http://www.journaldev.com/1997/servlet-jdbc-database-connection-example

[28] Wen, Mi, et al. "Secure data deduplication with reliable key management for dynamic updates in cpss." IEEE Transactions on Computational Social Systems 2.4 (2015)

[29] W. K. Ng, Y. Wen, and H. Zhu. Private data deduplication protocols in cloud storage. In S. Ossowski and P. Lecca, editors, Proceedings of the 27th Annual ACM Symposium on Applied Computing, pages 441–446. ACM, 2012.

[30] 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.

[31] M. Bellare, S. Keelveedhi, and T. Ristenpart. Message-locked encryption and secure deduplication. In EUROCRYPT, pages 296– 312, 2013.

[32] M. Bellare, S. Keelveedhi, and T. Ristenpart. Dupless: Serveraide encryption for deduplicated storage. In USENIX Security Symposium, 2013.

[33] Kevin D. Bowers, Ari Juels, and Alina Oprea. Hail: a high-availability and integrity layer for cloud storage. In Proceedings of the 16th ACM conference on Computer and communications security, CCS '09, pages 187–198, New York, NY, USA, 2009. ACM.

[34] Ari Juels and Burton S. Kaliski, Jr. Pors: proofs of retrievability for large files. In Proceedings of the 14th ACM conference on Computer and communications security, CCS '07, pages 584–597, New York,NY, USA, 2007. ACM.